

Merge, Project, and Bundle

Peter Svenonius
CASTL, University of Tromsø

August 7, 2012

Abstract

Chomsky has persuasively argued that there is a combinatory mechanism (which he calls Merge) which is central to the computational system of natural language. However, there is some unclarity regarding its scope. For example, does Merge combine the verb with aspect, or is there some other way for such features to be combined? There are combinations of features, such as the first person with the present tense in the English auxiliary *am*, for which Merge does not seem to be the optimal device. Using the term Merge for all feature combinations may reduce the content of the notion. In this paper, I outline a model for the generation of feature combinations which does not rely on Merge.

I suggest that the mechanism which creates extended projections and features bundles can be modeled as a finite state network. Transition represents the addition or projection of features, and states represents categories with features. Output states correspond to categories, while non-output states correspond to noncategorical features.

I accept Chomsky's (1957) arguments that such a system cannot replace Merge, which is necessary for embedding one extended projection inside another, but I suggest that it can account for the construction of extended projections and feature bundles. In this way, Merge can be restricted to those cases where it is truly needed, and can be studied in a sharper light. For example, I suggest that the chronic problem of labeling may be resolved if extended projections are not created by Merge.

1 Introduction

How many combinatoric systems are there in natural language? There has recently been a great deal of focus on the importance of Merge, the basic computational process posited by Chomsky (1995) that combines two syntactic objects to produce a new syntactic object (Hauser et al. 2002 and the lively debate it engendered). But in addition to syntactically complex structures, natural language has morphologically complex words and featurally complex morphemes and heads, as well as various levels of complexity in phonology.

Even if phonology and morphology are set aside, there are apparently two kinds of complex objects in syntax: heads and phrases. Phrases are the output of Merge. But heads are ordinarily also understood to be featurally complex, without being the output of Merge. How do features come together in featurally complex heads? What structure do they have?

One hypothesis which is sometimes entertained is that heads are not featurally complex. That is, syntactic features are syntactic atoms, so that each head bears only one feature (Kayne 2005). However, there are various indications that heads can be featurally complex. I assume that a syntactic feature is by definition a property relevant for syntax which is borne by a head or class of heads (see Svenonius 2007 for some discussion).

A head can require a specifier, like English T or German C, or not, like English C or German T (or even more clearly, Irish, McCloskey 1996). The property of attracting a specifier is a feature, by definition, and the properties of dominating T and being dominated by C are also syntactic features. Thus, such examples show that a head can combine two syntactic features. So at the very least we need to distinguish Merge from something else like “bundling”

whereby the positional property of dominating T is “bundled” with the dependency-forming property of taking a specifier, in a V2 language. The picture does not change significantly if T or C are decomposed (Pollock 1989, Rizzi 1997); a decomposed T or C reduces the number of combinations of features which must coexist on a single head, but cross-linguistic variation in subject movement, topic movement, focus movement, wh-movement and so on shows that features are combined in different ways in different languages.

Chomsky (2007; 2008) proposes a feature, which he calls the Edge feature, which is borne by syntactic objects which are subject to merge. Thus, every syntactically visible category bears an Edge feature. Features which are not themselves subject to Merge do not bear Edge features.

There are additional constraints on what merges with what. Lexical heads are usually assumed to have subcategorization features (see Emonds 2000 for discussion, including criticism of attempts to eliminate syntactic subcategorization). In addition, there is a functional sequence which constrains extended projections (Cinque 1999, Adger 2010a).

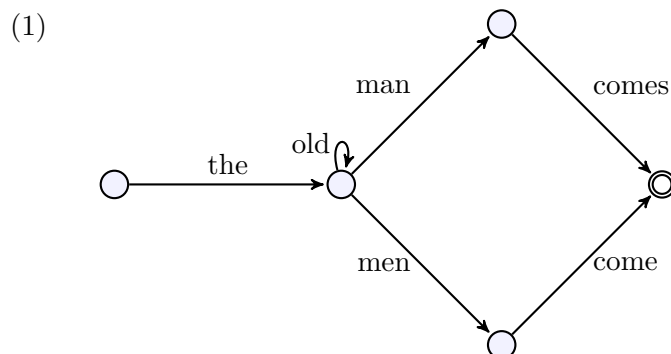
As for the structure of feature bundles in individual heads, it is sometimes assumed that these are unstructured sets, but there appear to be implicational relations among the features, which have been used to argue for “feature geometries” (e.g. Bonet 1991, Noyer 1992, Harley and Ritter 2002).

In this paper, I support the argument that there is a structure to syntactic features, and I suggest that the feature geometry is the same as the functional sequence, which I model as a set of finite state transition networks. The Edge feature distinguishes those parts of the functional sequence which are visible to Merge, which correspond to end states in the finite state network.

Thus, features must be combined by one of two mechanisms. One is the finite state network, which allows feature bundling in featurally complex heads. The other is Merge, which allows embedding of one category inside another, in complementation, specifier selection, and adjunction. When these mechanisms are combined, the result is extended projection building. Thus there is a basic difference between embedding and extended projection building, beyond the finer-grained differences among embedded complements, specifiers, and adjuncts.

2 Finite state transition networks

Finite state transition networks can produce language-like strings, capturing simple grammatical dependencies and can even produce infinite possibilities from finite means, as illustrated by the toy grammar in (1), from Chomsky (1957:19).

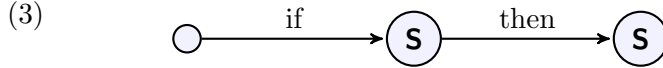


However, it is clear that finite state machines are inadequate for modeling natural language syntax (Chomsky 1955; 1956; 1957). As Chomsky observed, this can easily be seen from the fact that natural languages readily produce structures in which dependent components are separated from each other by arbitrary stretches of structure, for example in conditionals of the form *if* S_1 , *then* S_2 , where the occurrence of the antecedent *if* conditions a consequent clause (optionally introduced by *then*), but the two are separated by an arbitrarily complex S_1 , as in

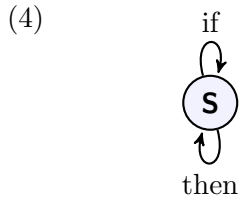
(2a). Similarly for structures of the form *Either S₃, or S₄*, as in (2b).

- (2) a. If you come to the party then [if I have records you like then you can play them].
- b. Either you come to the party or I'll [either have to play the radio or ask somebody else to DJ].

A finite state network has no external memory. The only mechanism is the transition from one state to another. If the dependency is modeled as in (3), then the grammar of S has to be repeated in each of two states, and the *if-then* rule will have to be repeated in each of them as well.



Loops are needed to allow the same possibilities to be reused. But loops like those in (4) will not work, because there is nothing to ensure that each time a clause is marked as an antecedent with *if*, there must be a consequence clause.



On the basis of such observations, Chomsky (1957:24) rejects finite state networks as a model of grammar, but comments: “The conception of grammar which has just been rejected represents in a way the minimal linguistic theory that merits serious consideration. A finite state grammar is the simplest type of grammar which, with a finite amount of apparatus, can generate an infinite number of sentences.”

The fatal flaw of the finite state machine is that it has no memory of how it got to whatever state it finds itself in at a given time. A phrase structure grammar separates the expansion process from the phrase structure rules licensing the process (and a transformational grammar has additional mechanisms), so Merge or its equivalent is necessary for natural language. But in order to fully understand the nature of Merge, we must understand what its purview is.

3 Extended projections

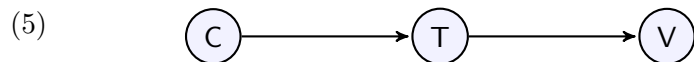
There has been quite a bit of discussion of what level of computing power natural language grammar must or should have (e.g. Vijay-Shanker and Weir 1994), and although much of the discussion is based on formal properties of strings, rather than the structures which are relevant for natural language, it remains clear that finite state networks are inadequate for modeling human language, as Chomsky pointed out over fifty years ago (Shieber 1985).

However, there are parts of natural language systems which seem to be rather like what might be expected from the output of a finite-state system. Word structure is a salient example (cf. Koskeniemi 1985).¹ Word structure is closely related to extended projections (witness the various models of word formation in which head movement plays a central role, and is constrained by the head-feature convention to complement sequences, e.g. Baker 1988, Halle and Marantz 1994, Julien 2002).

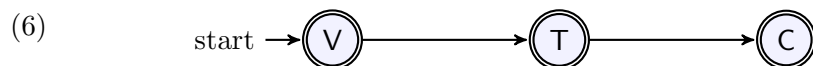
¹There are occasionally word-formation processes which seem to show greater than finite state complexity, e.g. the ‘loop’ in Turkish nominal morphology discussed in Hankamer (2004). Hankamer argues that the loop in question involves embedding of one extended projection inside another and hence, on my assumptions, would require Merge. Thus, words do contain major categorial boundaries, but the bulk of word formation and inflection are internal to extended projections.

Extended projection is Grimshaw’s (2005) term for the functional categories which belong with a given lexical head, namely C–T–V and P–D–N. These categories might host specifiers and adjuncts, and the lexical head at the bottom might have an embedded complement, but the extended projection itself is usually taken to be a total ordering on a relatively small set of features. Note that there are at least two functional sequences, one for V and one for N; separate sequences are sometimes posited for other categories. For example, den Dikken (2010) argues specifically that P is not in the functional projection of N, contra Grimshaw, but has its own extended projection.

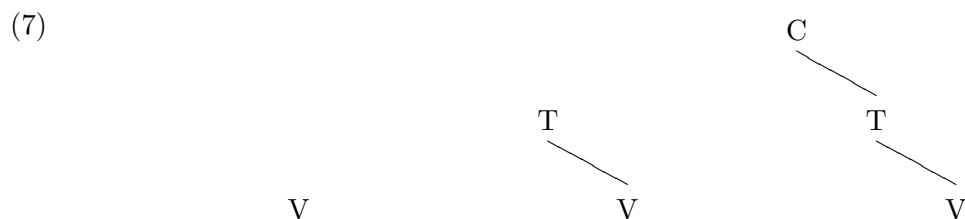
An extended projection, stripped of specifiers, adjuncts, and embedded complements, is well within the computational power of a finite state network.



For historical reasons, these diagrams have been drawn from the perspective of the parser, from top to bottom. From the perspective of production, what we want is a machine that generates syntactic structures from the bottom up. For example, with start state (root) and accepting states (Edge features) marked, as in the following.

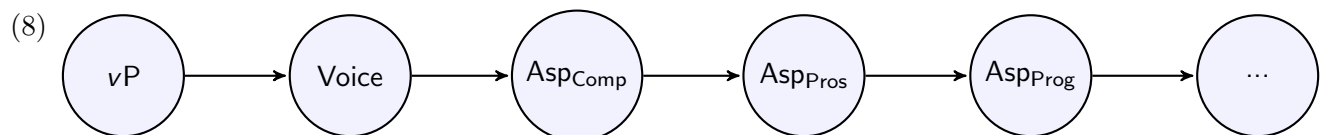


The projection machine in (6) generates the three trees in (7), and nothing else.



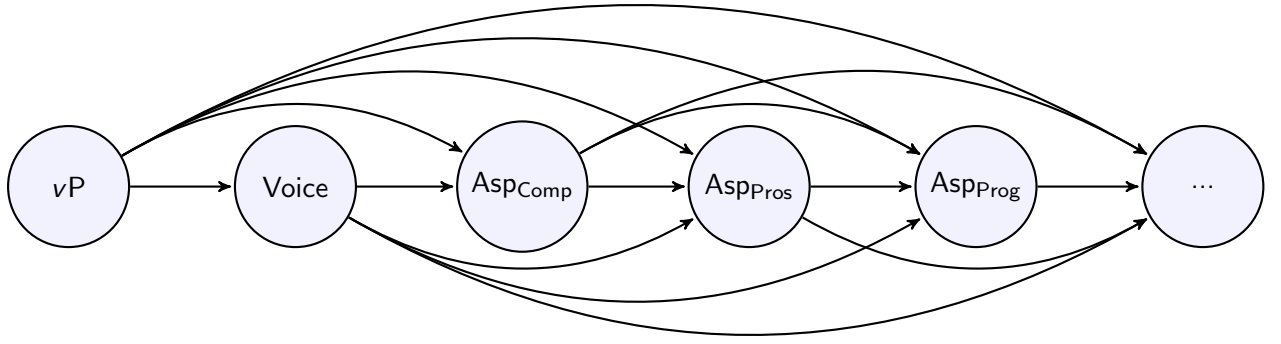
Following Chomsky (1957), a distinct mechanism is required for the embedding of one extended projection inside another (phrase structure rules and generalized transformations, in Chomsky 1957, or Merge, as in Chomsky 1995 inter alia). But extended projections themselves do not require these distinct mechanisms. The more powerful mechanisms can clearly also produce extended projections, and that is why it is ordinarily assumed that they do. But I suggest that nonprojecting features require a mechanism distinct from Merge, and that that mechanism has characteristics resembling a finite state network. I will argue that the same mechanism is at work in the extended projection.

A rich functional sequence like that of Cinque (1999) is structurally identical to a simple one like that of Grimshaw (2005). This is illustrated with a small part of Cinque’s proposed hierarchy in (8).



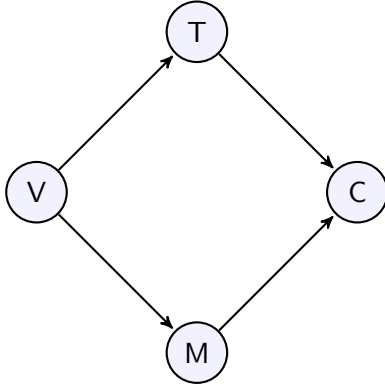
Cinque (1999) assumes that all functional nodes are present in every clause, but that nodes may have negative values corresponding to unmarked semantic interpretations (and often zero morphology). I discuss negative values of features below. However, if features are taken to be privative, then featural optionality is easily represented in the finite state network with multiple transitions from certain states, giving different options. For example, if all of the nodes from Voice through Asp_{Prog} are optional, this would be represented as in (9).

(9)



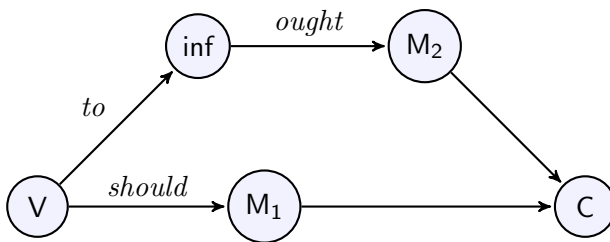
It is also straightforward to model complementary distribution in a finite-state system. For example, if the English clause includes either a modal or a finite tense, and these two represent distinct categories, this can be represented as follows.

(10) Complementary distribution



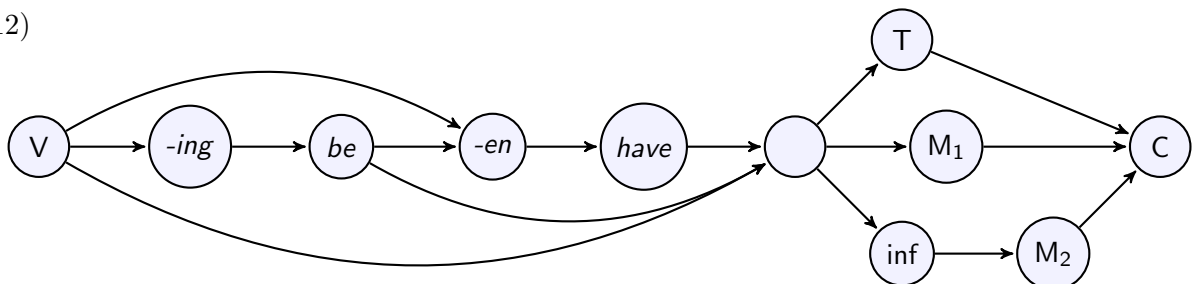
Similarly, implicational dependencies can be modeled. For example, if one class of modals, M_1 , combine directly with a bare infinitive, but another class of modals, M_2 , require the infinitival marker, this can be captured in a finite-state system.

(11) Implication



The combination of optionality, complementary distribution, and implicational dependency is sufficient to capture the dependencies observed in English “affix hopping” structures (Chomsky 1957), for example the correlation of verb form with auxiliary form.

(12)



It seems, then, that finite state transition networks might provide a useful way of thinking about extended projections. Merge (by whatever name) is still necessary for inserting specifiers and complements (of the traditional sort, i.e. those which are not part of the extended projection), but if extended projection formation is eliminated from the purview of Merge, it might turn out that Merge looks slightly different from what is often assumed, as I will argue below.

Adger (2011; to appear) has also proposed a formal distinction between extended projection formation and category embedding. On Adger’s proposal, extended projections are formed by self-merge, rather than external merge. By this he means that rather than functional material being introduced as a head, the extended projection merges with itself, creating a new projection, the label of which is provided by a labeling algorithm. If that proposal were adopted, then the system that I am proposing here could be thought of as the theory of the labeling algorithm.

4 Extensions

4.1 Transfer and Tucking

Chomsky’s (1993) proposed Extension Condition holds that structure building occurs at the root of the tree. In Chomsky’s terms, Merge always “extends its target.”² It is a plausible and natural condition on Merge and derives a strict form of cyclicity. However, there have been several proposals which require it to be weakened in different ways. For example, Richards (2001) proposes that when multiple specifiers are attracted to a single head, the first adheres to the Extension Condition, but subsequent attractees ‘tuck’ under the first. This means that segment formation, as opposed to the projection of a first specifier, falls outside the Extension Condition. Richards accordingly revises the Extension Condition as a condition on feature checking, rather than on Merge.

More radically, Chomsky (2007; 2008) proposes that non-phase heads cannot attract specifiers until they receive the appropriate features from a phase head, by an operation he calls Transfer. This would appear to require the abandonment of the Extension Condition as originally conceived. This means that the only guarantor of cyclicity seems to be the phase; Merge can insert phrases into non-phases, or incomplete phases.

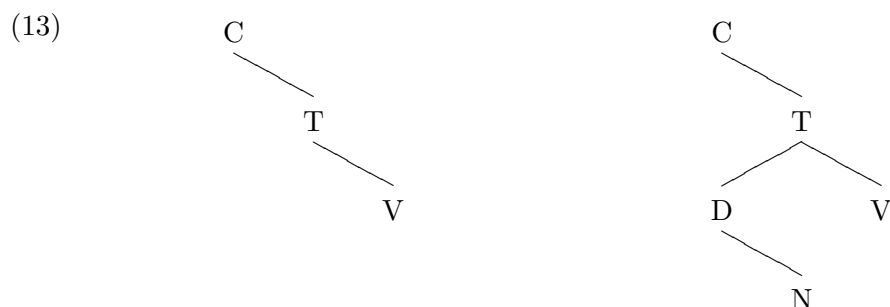
For example, suppose for the sake of argument that an adverb is not by itself a phase. The original Extension Condition would allow Merge to target an unembedded adverb (as in the construction of *very often*), but would not allow movement into the projection of an adverb which had already been merged into a tree. Nor would Tucking, as proposed by Richards, since he assumes that specifier-forming features on a head must be satisfied before the projection of that head is embedded. These seem to be good results. But Transfer opens up the possibility anew, because it is not clear what would prevent a phase head C from transferring its specifier-forming features to an adverb.

However, if the construction of extended projections is distinct from Merge, then the Extension Condition can be rethought. Suppose that the Extension Condition is the requirement that Merge increase its target, where its target is not an unembedded phrase but an unembedded extended projection (a syntactic object with an Edge feature), and ‘increase’ means not necessarily adding to the top, but adding a specifier-dependency somewhere in the projection line. Apparent violations of the Extension Condition as proposed by Richards and Chomsky involve specifiers merging lower in an extended projection than expected. If Transfer never targets phrases which are already embedded as specifiers in the extended projection (as in the

²Internal Merge forms a dependency between a higher position and a lower one. The “target” in this sense is the higher position, the one which gains a specifier and hence, on usual conceptions, is “increased” by the addition of a segment. If we think of Internal Merge simply as the forming of an asymmetric dependency, where the higher position is superordinate to the lower one, then the Extension Condition means that the superordinate position must be unembedded at the time the dependency is formed.

hypothetical case of the adverb), then this could be because Merge cannot apply to embedded projections, only to unembedded extended projections. Within the boundaries of a phase, a span of an extended projection can be taken to be transparent to Merge.

In other words, if Transfer is correct, then Merge can operate on the lefthand tree in (13) (from (7) above, projected by the machine in (6)) to produce the righthand tree. As a graphic convention, I draw phrases inserted by Merge under branches which slope down to the left, and phrases which are added by Project above branches which slope down to the right (this is based on Brody 2000a;b, who argues that there is no distinction between X and XP, and develops a theory of the distinction between heads and specifiers to which the approach here is amenable).

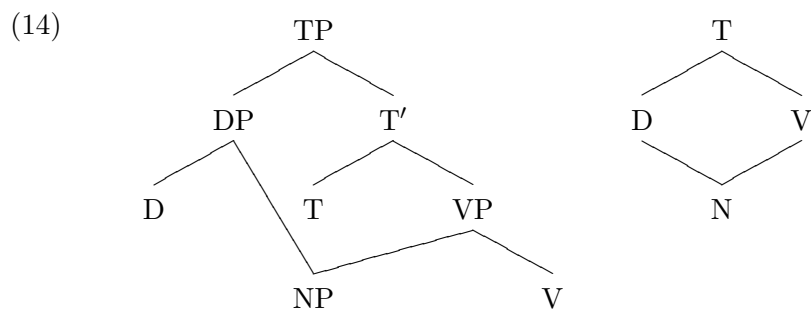


What Merge cannot do, by hypothesis, is to extend the embedded D projection in (13) by adding specifiers to it. These must be added before D is embedded in a distinct extended projection.

I am suggesting a split between Merge and Projection. Projection is responsible for the construction of extended projections, including those which are not phases (such as adverbs, by hypothesis) and those which are (such as the *vP*). The idea is that Merge is necessary for embedding one category inside another, while Project allows the construction of complex categories. The Project system may also be responsible for labeling (essentially as in Adger 2011). The clause normally consists of two phases, on phase theory, but one extended projection, so there is a question whether Merge or Project is responsible for embedding *vP* under T (or Asp, or Voice, or whatever). I suggest below that Project, rather than Merge, combines a *vP* with the C/T-domain, and similarly for *nP* and the P/D-domain, but it is possible that this too is actually a matter of Merge. I develop the split between Merge and Project somewhat more in the following subsection.

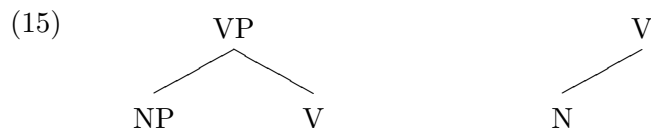
4.2 Sideward movement and Banyan trees

There are various indications that what is merged inside VP is not the entire DP argument, but a subpart of it; in other words, only the noun is interpreted as part of a VP (an event description), while a DP is interpreted as part of the clause (Haeberli 2002, Williams 2003, Sportiche 2005, Svenonius 2005). How this subpart, suppose we call it NP, grows into a DP has not been entirely clear. What seems to be needed is a structure like the following. The diagram on the left has traditional heads and projection and uses multidominance to represent movement (hence the structure is not technically a tree but is equivalent to one once the multidominance is resolved). The equivalent Brodian structure on the right abandons the head-phrase distinction and uses branches sloping down to the right for extended projection complementation, branches sloping down to the left for specifiers.

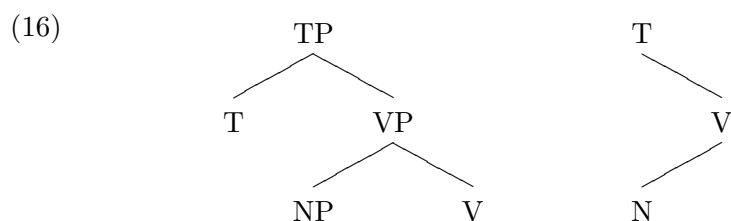


This is the equivalent of movement into a complement position (sideward movement, in the sense of Nunes 2004), another kind of violation of the Extension Condition, and structurally similar to the hypothetical case of moving into the specifier position of an adverbial modifier, which is unattested. Consider, however, how the interaction of Merge and Project along the lines discussed above can distinguish the case in (14), which I take to exist, from the case of movement into a specifier position, which I take not to.

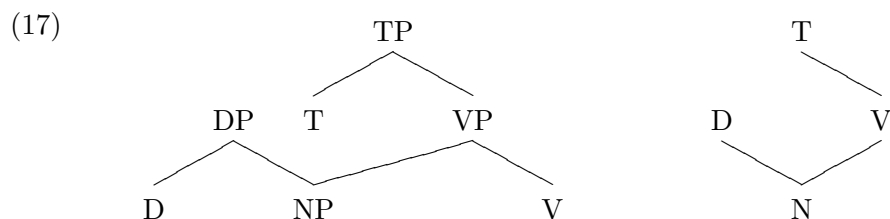
V and NP are distinct extended projections, which I assume means that they can be combined by Merge. I continue to provide the equivalent traditional and Brodian trees.



V projects T, in a way that gives the effect usually assumed to involve the merge of T with VP, as in the traditional tree on the left, but which I now assume is effected by a distinct mechanism, a projection engine like the one in (6), more perspicuously illustrated by the notational variant on the right.

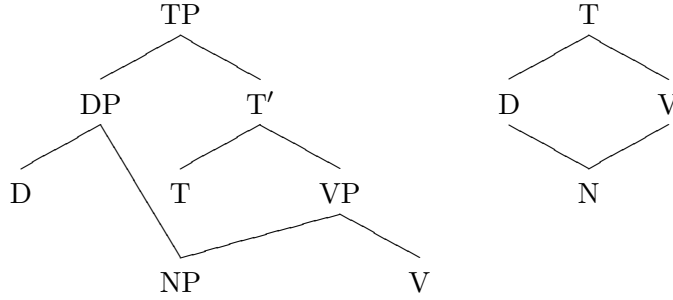


Now if Project is not constrained by the Extension Condition (which, it will be recalled, ensures that Merge only adds dependencies from unembedded heads), then N can project D even when N is dominated by V, as illustrated in (17).



Now, Merge recombines the two distinct phrases TP and DP, in compliance with the Extension Condition, deriving the structures originally presented in (14).

(18)



This allows a restricted kind of sideward movement. The hypothetical movement into the specifier of an adverb would still not be permitted; the adverb could only increase its own extended projection, through Projection, but could not absorb a target of movement, because the added projections would not c-command anything in the main projection line. When an embedded phrase adds projections, what results is a “Banyan tree,” with multiple roots.³

An interface condition can be assumed to require that each phase which is interpreted has a single root. Thus, the offshoot projection may acquire specifiers or modifiers not included in NP, but whatever dependents it contains must ultimately be dominated by a projection of the verb, as in (14).

So in the hypothetical case where an adverb attracts a specifier by movement, the target of movement would have to be a projection of the adverb, and so could not c-command anything in the main projection line. Once the extended adverbial projection is remerged back into the structure, by Merge, the Extension Condition will then prevent any further Merging of lower material.

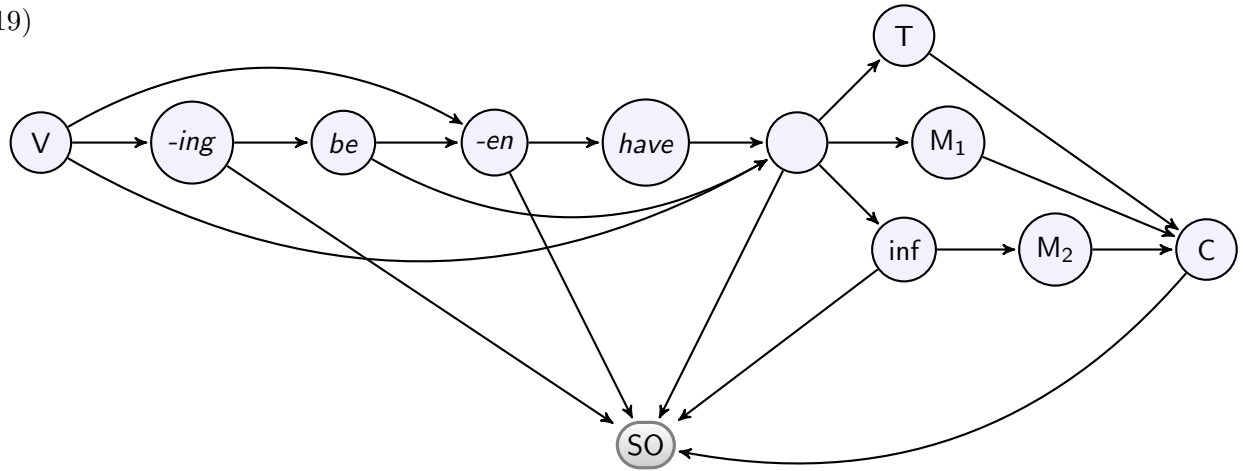
5 Feature structures

I have argued for a basic distinction between Merge, which combines extended projections with each other, and Project, which allows extended projections to be built. Project, I have suggested, has computational properties similar to those of a finite-state machine. Arguably, this is the right level of complexity for feature bundles as well. Adger (2010b) discusses the complexity which is made possible by GPSG and HPSG feature systems, and suggests that a Minimalist model must avoid them in order to avoid redundancy with Merge as a structure-building operation. He proposes the *No Complex Values Hypothesis* (NCV) for feature structures, to restrict the embedding function to Merge (see also Adger and Svenonius 2011). Hence, feature bundles, on Adger’s view, must be constructed by something other than Merge. What I suggest here is a framework for understanding how features bundles can be constructed, one which allows something like the NCV to fall out from the architecture.

Take, for example, the Edge feature, which Chomsky has proposed distinguishes those objects which are visible to Merge from those which are not. I have suggested that it can be modeled as an “accepting state” in a finite state network. This can be modeled in terms of a feature (a second-order feature, in the sense proposed by Adger and Svenonius 2011) being added to certain nodes, as in the following modification of the affix-hopping network from (12).

³Adopting a trace/copy notation, Bobaljik and Brown (1997) call the relation represented in (17)–(18) ‘inter-arboreal,’ while Nunes (2001; 2004) calls it ‘sideward movement.’ Wilder (1999) calls such structures ‘multiple dominance’ structures, and Citko (2005) and Bachrach and Katzir (2009) call them ‘parallel merge’ structures. All of the proposals just mentioned assume that such structures are ultimately embedded under a single root node, as do I. I call the structures banyan trees, following Svenonius (2004) to distinguish them from the notion of multidominance as an alternative model of movement as remerge or internal merge (on which see Starke 2001, Gärtner 2002) and from the notion of merge in parallel workspaces which is normally assumed for specifiers (on which see Uriagereka 1999, Narita 2011, Lohndal 2012).

(19)

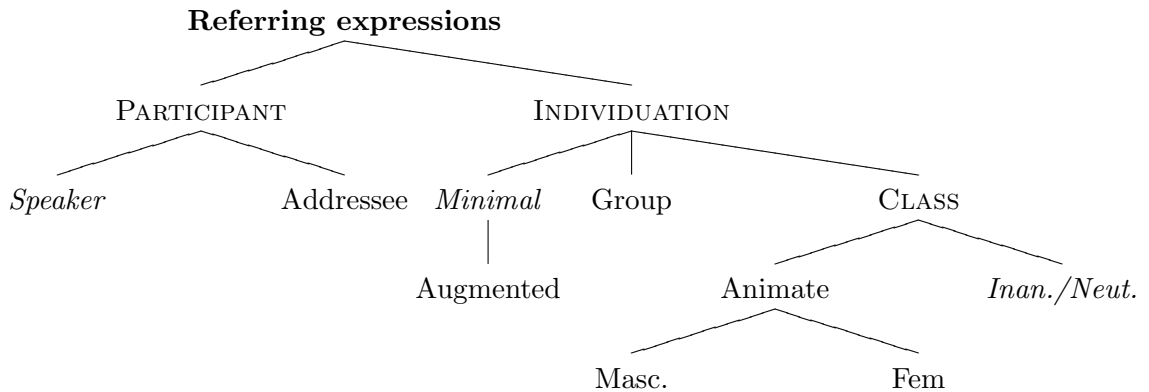


In this network, clauses, bare verbs (the blank node), *to*-infinitives, and present and past participles can be marked with the Edge feature, making them visible to Merge and thus potentially embeddable in distinct extended projections. Other categories cannot. Merge operates on SO's (syntactic objects). The double-edged ring notation for accepting states can be seen as a notational variant of this.

5.1 Φ -features

Work on morphosyntactic features has observed a number of implicational dependencies, for example Bonet (1991), Noyer (1992) and Harley and Ritter (2002). The latter propose a feature “geometry” for ϕ in the form of the following tree, which is intended to subsume (in the sense of Shieber 1985) the ϕ -feature systems found in different languages.

(20)



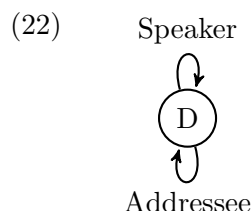
The tree is intended to capture implicational dependencies (small capitals indicate major categories, and italics indicate default values). There are implications among the larger categories of person, number and gender, for example if a category is large enough to distinguish first and second person features (Participant) from third, then it is a referring expression, and can distinguish number and gender as well (under Individuation). However, Individuation is not dependent on Participant, so a category can carry Individuation features (gender and number) without carrying Participant features (person), as in participle agreement. There are also implications within the categories, for example Masculine and Feminine are dependent on Animate, so they imply it; any noun which is specified as Masculine or Feminine is also Animate. The dependence of Augmented on Minimal is how Harley and Ritter achieve the implication that if a language has a dual, it has a plural. I will briefly discuss person, number, and gender in turn, showing how a structure like that in (20) could be understood both as a part of linguistic competence and as an output of that competence system.

5.2 Person

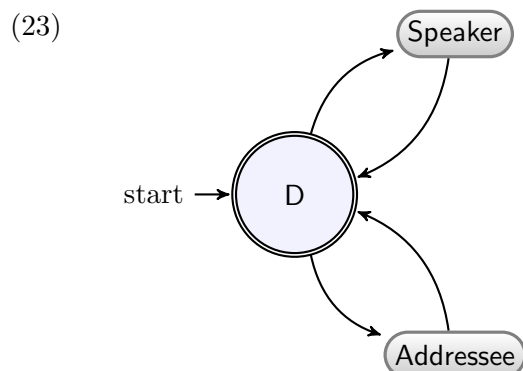
For person, Harley and Ritter (building on Noyer 1992) suggest that third person involves just the Referring Expression node, without a Participant feature. For present purposes I will assume this node is D. First and second person are distinguished by two features, Speaker and Addressee. In systems that make a distinction between inclusive and exclusive first person, Speaker without Addressee is exclusive, and Speaker with Addressee is inclusive (setting aside Harley and Ritter’s Participant node for the moment).

- (21)
- a. D (third)
 - b. D, Addressee (second)
 - c. D, Speaker (first exclusive)
 - d. D, Speaker, Addressee (first inclusive)

Assuming that iteration of features adds no meaning, the network in (22) produces all and only the feature combinations in (21).

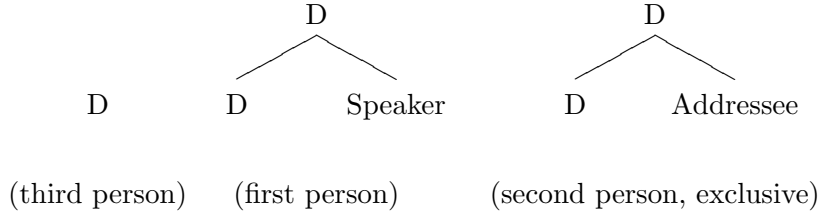


More explicitly, this can be represented as in (23), including nodes for the feature labels and marking the category as an accepting state (i.e., being able to host an Edge feature) to make explicit the difference between the category D, which determines distribution, in the Bloomfieldian sense, and the person features, which have syntactic relevance but do not project categories.



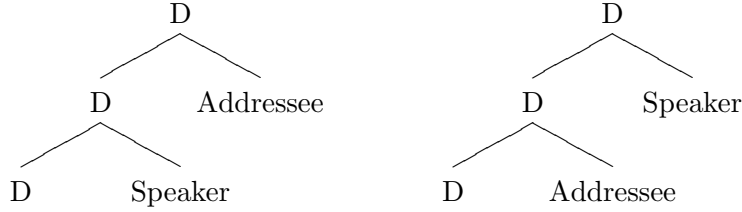
In the creation of extended projections, I assumed that each transition led to the projection of dominating structure. If we maintain that assumption then something must be said about how loops like those in (23) should be interpreted. I suggest that they can be interpreted in a way similar to adjunction. The transition from D to Speaker, for example, would be expected to project a Speaker node above D, but the transition back to D projects the D node above the Speaker node. If the identity of the D node is maintained, then this leads to a situation where one segment of D dominates Speaker and the other does not, which is the nature of an adjunction structure. Thus, I assume that the network in (23) generates trees like the following for third, first, and second person exclusive respectively.

(24)



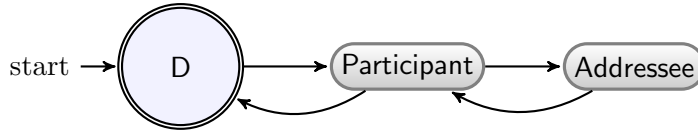
The independence of the Speaker and Addressee loops allows two additional possibilities, which I assume are equivalent, by commutativity, both meaning first person inclusive.

(25)



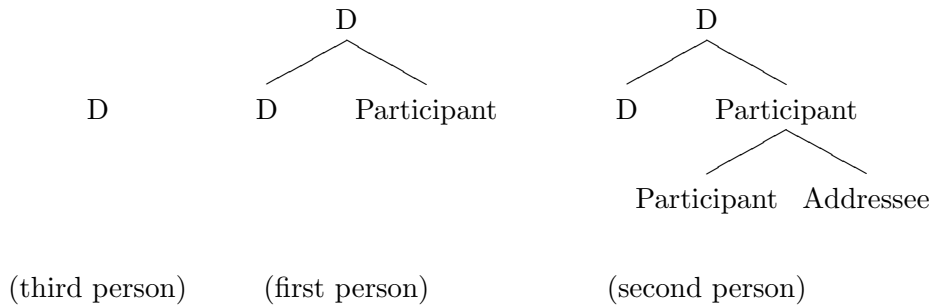
Languages which do not make an inclusive-exclusive distinction might have a slightly different structure for person, for example with addressee dependent on a ‘participant’ node with a default interpretation of speaker in the absence of the Addressee feature. This would be generated by the following network.⁴

(26)



This would generate all and only the following trees.

(27)

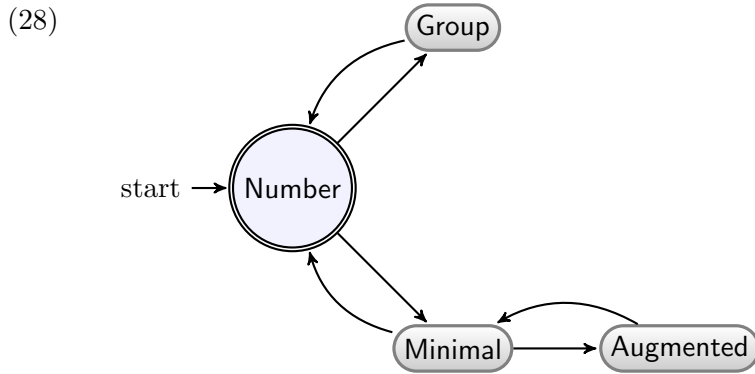


5.3 Number

Crosslinguistic variation is somewhat limited in person systems. There is a greater degree of crosslinguistic variation in number systems. See Noyer (1992), Harbour (2007) for discussion. The structure proposed by Harley and Ritter has one pair of features in an asymmetric implicational relation, where Augmented implies Minimal, and another feature which is independent of them, namely Group.

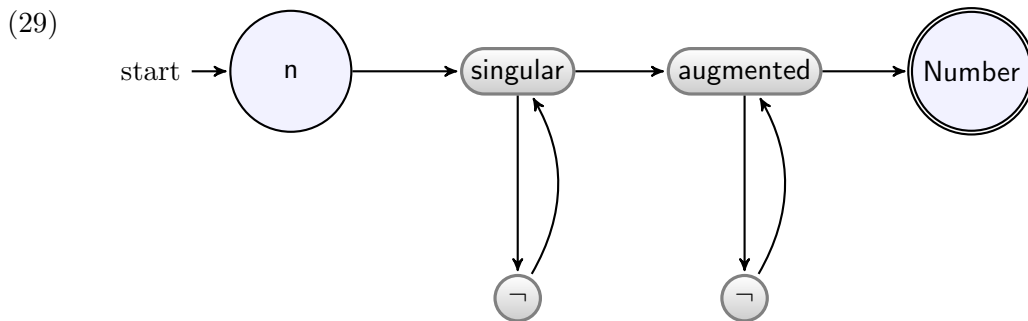
⁴I am assuming that different person networks can be learned, an assumption which is necessary for number and gender, discussed below. A stronger universalist position for person (like that of Harley and Ritter) would try to get both types of person system (with and without clusivity) from a single network. A single network can generate both the lefthand tree in (25) and the rightmost tree in (27), as well as all the simpler trees, and nothing else, if a transition is added from Addressee to D in (26). Essentially, this allows Addressee to be generated under Participant, and then either take the option of tracking back through Participant (to give second person) or “moving” up to be a direct dependent of D (to give first person inclusive). Having the clusivity contrast would then correspond to a single transition, but more would have to be understood about learning mechanisms and constraints on feature networks before this could be a serious proposal.

This can be modeled with the following structure.



This represents a relatively complex system. Many languages make use of only a subset of these features. A language with a simple singular-plural distinction, only has a contrast between having Group and not having it (Harley and Ritter assume that in such languages, Minimal is supplied as a default if Group is not present). Languages with a three-way distinction between singular, dual, and plural need one additional feature, in the privative system assumed by Harley and Ritter, and they assume that Minimal and Group are both present in the dual. The Augmented feature is necessary for languages which additionally have a paucal number.

Harbour (2007; 2011a;b) develops a somewhat different theory of number, which is based on obligatory binary-valued features singular and augmented (which he relates to atomicity and cumulativity in aspectual systems). Some languages just make use of one or the other, but if a language makes use of a feature, then it is obligatorily present, with or without negation. A language with a three-way distinction between singular, dual, and plural makes use of both features. The combinations of features that Harbour proposes are generated by the following network.



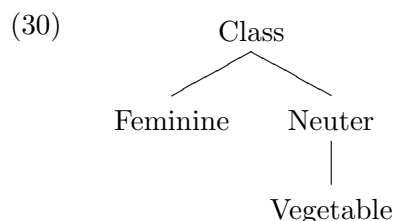
This generates the combinations [singular,–augmented] (singular), [–singular,–augmented] (dual), [–singular,augmented] (plural), and [singular,augmented] (which Harbour argues is uninterpretable and hence filtered out if generated).

Harbour's theory is also distinct in that it treats features as functions, with the result that multiple instantiations of a feature on the same head are not necessarily redundant; in particular he argues that [–augmented] can apply to the output of [augmented] without contradiction. In this way he derives the trial, and suggest that languages which have a trial allow augmented to iterate, which can be achieved in this model by adding a loop on the augmented back to itself, without passing through negation.⁵

⁵Note that Harbour's model means that commutativity of feature combinations cannot be taken for granted, contra what I assumed above for person.

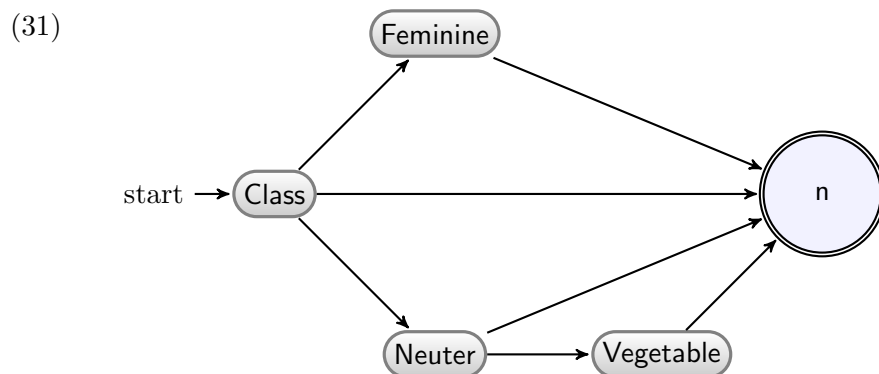
5.4 Gender

To an even greater extent than with number, there is interesting crosslinguistic variation in the structure of gender (see Corbett 1991 for an overview). Pensalfini (2003) discusses Jingulu, where a four-way distinction can be made among masculine, feminine, neuter, and vegetable.⁶ Gender agreement reveals a hierarchy of defaults, with feminine and neuter nouns sometimes controlling masculine agreement, and vegetable class nouns sometimes controlling neuter. This suggests that vegetable is formally dependent on neuter, and that feminine and neuter are formally dependent on ‘masculine,’ which can be identified as a kind of default and can be renamed “Class” for comparison with the Harley and Ritter (2002) tree, as represented in the following diagram (lightly adapted from Pensalfini 2003:169; compare also Corbett 2006:152, where a node is added between Class and Feminine, corresponding to Harley & Ritter’s ‘Animate’).



The idea here is that masculine gender nouns simply have a Class, while feminine nouns additionally have a Feminine feature and neuter nouns additionally have a Neuter feature; vegetable class nouns have both Class and Neuter. In partial agreement, one or more of the more marked nodes in the tree is not copied, so that vegetable can become neuter or masculine, and neuter and feminine can become masculine.

It is unclear exactly how to model optional partial agreement. One possibility is that it is related to optional pied-piping, where a target for movement optionally carries along dominating structure or not. If that is the case, then the cladistic tree in (30) is inverted compared to the syntactic tree of features, because it is Class which is always attracted by an agreement probe (yielding at least masculine agreement), and it optionally pied-pipes the other material, which must then be higher (see Chomsky 1995 on agreement as feature movement not pied-piping a category). This would be generated by a network like the following.



The lexicon contains nouns which have non-arbitrary distributions of these features,⁷ and perfect matching can be assumed for root insertion, so that only a masculine noun can be inserted in tree that lacks Neuter or Feminine features. Agreement copies feature structures in

⁶Pensalfini describes the distinction as essentially lexical but partially semantically based, like many gender systems; e.g. the ‘vegetable’ category includes many long thin things, like penises, didgeridoos, gullies and rainbows. I have no reason to doubt Pensalfini’s characterization, but see Plaster and Polinsky (2010) for a cautionary tale from Dyirbal concerning the limits of such semantic generalizations.

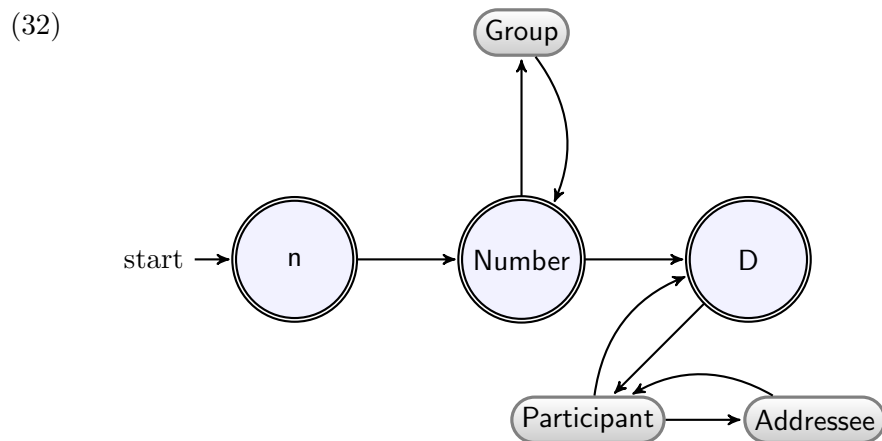
⁷Which I assume is because the syntactic features in lexical entries are stored outputs of the same grammar which produces trees online, as proposed by Starke (2011).

a way that preserves the dependencies. A probe for *n* would carry along with it all agreement features, but in Jingulu, agreement probes target Class, which only optionally pied-pipes the other features.⁸

On Cable’s (2007) theory of pied-piping, optionality of pied-piping has to do with optional placement of the goal of probing. This could be accommodated in the present discussion by assuming that in Jingulu, there are additional transitions back to Class from the various gender feature nodes. When the forward transitions are followed, then Class is low in the structure generated, and this leads to partial agreement. When the backward transitions are used, then adjunction structures are formed, and Class is high in the structure, leading to full agreement. However, this potentially complicates the matter of matching nouns to the structures generated. There are several other plausible ways of treating optional partial agreement, so I will not dwell on the matter further.

5.5 Illustrations

Combining the above observations, we can sketch a model of the extended projection of *N* in a given language, taking English as an example first, in (32). English gender has no grammatical significance beyond the spell-out of pronouns,⁹ so I will assume it is not represented in the feature structure but is part of nonfeatural, conceptual semantics.



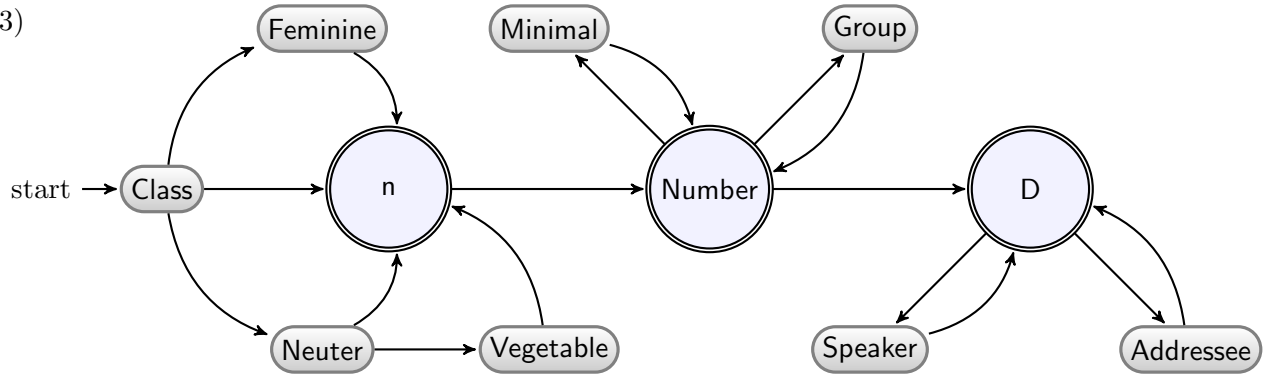
The category *n* is designated as an output state, for example in compounds, and *Number* is also an output state, on the assumption that there are determinerless NumberPs in English, for example measure phrases, or perhaps indefinites. English makes a simple singular-plural distinction, with no dual or paucal, so the *Group* feature is the only one present for number, and there is no inclusive-exclusive distinction, so *Addressee* is dependent on *Participant*. I set aside case, which I treat in a separate paper.

A more complex example is provided by Jingulu in (33). The gender system is at the bottom of the nominal extended projection, and there is obligatorily at least one gender feature on each noun, relevant for gender agreement. Jingulu makes a three-way distinction among singular, dual, and plural, so it needs both the features *Minimal* and *Group*, in the Harley and Ritter system. *Number* can again be assumed to be an output state, for indefinites. After that, there is *D*, for definite and pronouns, and in the case of first and second person pronouns Jingulu makes the inclusive-exclusive distinction, so has the independent *Speaker* and *Addressee* configuration. Again, I set aside case.

⁸Alternatively, perfect matching is not enforced for roots, so that any kind of noun can be inserted into an ‘underspecified’ Class node, and agreement is always complete.

⁹Unless the matching requirement in anaphor binding is grammaticized, i.e. in sentences like *She bit herself*/**himself*; sentences like *My dog bit herself* suggest it is not, and can be handled in the same way as cross-sentential anaphora.

(33)



A natural distinction appears to emerge between what are typically assumed to be categories (and can stand alone as projections) and what are typically assumed to be features (and are dependent on a category), namely categories are generalizations about relatively closed featural systems. Essentially, each category is a micromodule. This begins to suggest how a functional sequence of categories might emerge from the interaction between general cognitive systems and common experience, rather than being ‘hardwired.’

6 Conclusion

I have suggested a model of the functional sequence and of feature structures which lessens the burden on Merge as the sole combinatory system of natural language. For example, the problem of labeling the structures produced by Merge, as discussed in Collins (1997) and Boeckx (2008) among others, looks quite different if Project is responsible for labeling (cf. Adger 2011). I also believe this proposal relieves some of the tension between the Cartographic program of Cinque and Rizzi and Chomsky’s Minimalist program. Certain assumptions about locality and feature transfer are problematic when scaled from a Minimalist tree consisting of C-T-*v*-V to a Cinquean tree with thirty heads corresponding to T. For example, the proposal by Epstein and Seely (2002) connecting cyclicity directly to Merge is problematic in a Cinquean tree, but when the Cinquean projections are added by Project, rather than by Merge, the relationship of Merge to the phase becomes more tractable. Furthermore, Cinque’s strong universalist stance seemed hard to reconcile with the Minimalist effort to reduce UG to a bare minimum. Here, I suggest an architecture which makes sense of the relationship between categories and features, pushing forward the possibility that the functional sequence and the difference between bundled and projecting features could be emergent.

The proposal begs a theory of how the networks emerge. I assume that general cognitive constraints greatly limit the kinds of systems which do emerge. Some of those constraints correspond to properties of modules or faculties, like the one distinguishing discourse participants or the one distinguishing cumulative from noncumulative instances of substance or experience. Other constraints have to do with the computational system. Feldman (2006) proposes an algebraic theory of concept learning which makes it possible to distinguish more and less highly valued regularities in data patterns. Certain kinds of regularities, such as straightforward implications, are valued over other kinds, such as exclusive disjunctions. An initial attempt to apply this kind of theory of learning to the emergence of the functional sequence has been made in Svenonius (2010). Obviously much more work is necessary, but I hope that this paper makes some small contribution in that direction.

References

Adger, David. 2010a. A minimalist theory of feature structure. In *Features: Perspectives on a Key Notion in Linguistics*, edited by Anna Kibort and Greville G. Corbett, pp. 185–218.

- Oxford University Press, Oxford.
- Adger, David. 2010b. Variability and grammatical architecture. Ms. Queen Mary University of London; to appear in a volume based on Barcelona Workshop on Linguistic Variation in the Minimalist Framework.
- Adger, David. 2011. Labels and structures. Ms. Queen Mary University of London.
- Adger, David. to appear. *A Syntax of Substance*. MIT Press, Cambridge, Ma.
- Adger, David and Peter Svenonius. 2011. Features in minimalist syntax. In *Oxford Handbook of Linguistic Minimalism*, edited by Cedric Boeckx, pp. 27–51. Oxford University Press, Oxford.
- Bachrach, Asaf and Roni Katzir. 2009. Right node raising and delayed spellout. *InterPhases: Phase-theoretic investigations of linguistic interfaces* pp. 283–316.
- Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago.
- Bobaljik, Jonathan David and Samuel Brown. 1997. Interarboreal operations: Head movement and the extension requirement. *Linguistic Inquiry* 28 2: 345–356.
- Boeckx, Cedric. 2008. *Bare Syntax*. Oxford University Press, Oxford.
- Bonet, Eulalia. 1991. *Morphology after Syntax: Pronominal Clitics in Romance*. Ph.D. thesis, MIT, Cambridge, Ma.
- Brody, Michael. 2000a. Mirror Theory: Syntactic representation in Perfect Syntax. *Linguistic Inquiry* 31 1: 29–56.
- Brody, Michael. 2000b. Word order, restructuring, and Mirror Theory. In *The Derivation of VO and OV*, edited by Peter Svenonius, pp. 27–43. John Benjamins, Amsterdam.
- Cable, Seth. 2007. *The Grammar of Q: Q-Particles and the Nature of Wh-Fronting, as Revealed by the Wh-Questions of Tlingit..* Ph.D. thesis, MIT, Cambridge, Ma.
- Chomsky, Noam. 1955. The logical structure of linguistic theory. Published 1975 by Plenum Press, New York.
- Chomsky, Noam. 1956. Three models for the description of language. *IRE transactions on information theory* 2 3.
- Chomsky, Noam. 1957. *Syntactic Structures*. Mouton, 's-Gravenhage.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, edited by Kenneth Hale and Samuel Jay Keyser, pp. 1–52. MIT Press, Cambridge, Ma.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge, Ma.
- Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, edited by Uli Sauerland and Hans-Martin Gärtner, pp. 1–18. Mouton de Gruyter, Berlin.
- Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, edited by Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, pp. 133–166. MIT Press, Cambridge, Ma.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press, New York.
- Citko, Barbara. 2005. On the nature of merge: External merge, internal merge, and parallel merge. *Linguistic Inquiry* 36 4: 475–496.
- Collins, Chris. 1997. *Local economy*. No. 29 in Linguistic Inquiry Monographs. MIT Press, Cambridge, Ma.
- Corbett, Greville G. 1991. *Gender*. Cambridge University Press, Cambridge.
- Corbett, Greville G. 2006. *Agreement*. Cambridge University Press, Cambridge.
- den Dikken, Marcel. 2010. On the functional structure of locative and directional PPs. In *Mapping Spatial PPs: Cartography of Syntactic Structures, vol. 6*, edited by Guglielmo Cinque

- and Luigi Rizzi, pp. 74–126. Oxford University Press, Oxford.
- Emonds, Joseph. 2000. *Lexicon and Grammar: The English Syntacticon*. Mouton, Berlin.
- Epstein, Samuel David and T. Daniel Seely. 2002. Rule applications as cycles in a level-free syntax. In *Derivation and explanation in the minimalist program*, edited by Samuel David Epstein and T. Daniel Seely, pp. 65–89. Blackwell, Oxford.
- Feldman, Jacob. 2006. An algebra of human concept learning. *Journal of Mathematical Psychology* 50 4: 339–368.
- Gärtner, Hans-Martin. 2002. *Generalized Transformations and Beyond: Reflections on Minimalist Syntax*. Akademie Verlag, Berlin.
- Grimshaw, Jane. 2005. Extended projection. In *Words and Structure*, edited by Jane Grimshaw. CSLI, Stanford, Ca. Final version of 1991 ms.
- Haeberli, Eric. 2002. *Features, Categories, and the Syntax of A-Positions*. Kluwer, Dordrecht.
- Halle, Morris and Alec Marantz. 1994. Some key features of Distributed Morphology. *MIT Working Papers in Linguistics* 21: 275–288.
- Hankamer, Jorge. 2004. An ad-phrasal affix in Turkish. *MIT Working Papers in Linguistics* 46: 289–299.
- Harbour, Daniel. 2007. *Morphosemantic Number*. Springer, Dordrecht.
- Harbour, Daniel. 2011a. Descriptive and explanatory markedness. *Morphology* 21 2: 223–245.
- Harbour, Daniel. 2011b. Valence and atomic number. *Linguistic Inquiry* 42 4: 561–594.
- Harley, Heidi and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78 3: 482–526.
- Hauser, Marc D., N. Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–1579.
- Julien, Marit. 2002. *Syntactic Heads and Word Formation*. Oxford University Press, New York.
- Kayne, Richard S. 2005. Some notes on comparative syntax, with special reference to English and French. In *The Oxford Handbook of Comparative Syntax*, edited by Guglielmo Cinque and Richard S. Kayne, pp. 3–69. Oxford, New York.
- Koskeniemi, Kimmo. 1985. Compilation of automata from morphological two-level rules. In *Fifth Scandinavian Conference of Computational Linguistics*, pp. 143–149.
- Lohndal, Terje. 2012. *Without Specifiers: Phrase Structure and Events*. Ph.D. thesis, University of Maryland, College Park, Md.
- McCloskey, James. 1996. Subjects and subject positions in Irish. In *The Syntax of the Celtic Languages: A Comparative Perspective*, edited by Robert D. Borsley and Ian Roberts, pp. 241–83. Cambridge University Press, Cambridge.
- Narita, Hiroki. 2011. *Phasing in Full Interpretation*. Ph.D. thesis, Harvard University, Cambridge, Ma.
- Noyer, Ralf. 1992. *Features, Positions and Affixes in Autonomous Morphological Structure*. Ph.D. thesis, MIT, Cambridge, Ma.
- Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry* 32 2: 303–344.
- Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. MIT Press, Cambridge, Ma.
- Pensalfini, Robert. 2003. *A Grammar of Jingulu, an Aboriginal Language of the Northern Territory*. Pacific Linguistics, Canberra.
- Plaster, Keith and Maria Polinsky. 2010. Features in categorization, or a new look at an old problem. In *Features: Perspectives on a Key Notion in Linguistics*, edited by Anna Kibort and Greville G. Corbett, pp. 109–142. Oxford University Press, Oxford.
- Pollock, Jean Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20 3: 365–424.

- Richards, Norvin. 2001. *Movement in Language: Interactions and Architectures*. Oxford University Press, Oxford.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar: Handbook in Generative Syntax*, edited by Liliane Haegeman, pp. 281–337. Kluwer, Dordrecht.
- Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8 3: 333–343.
- Sportiche, Dominique. 2005. Reconstruction, binding and scope. In *The Syntax Companion*, edited by Martin Everaert and Henk van Riemsdijk, vol. IV. Blackwell, Oxford.
- Starke, Michal. 2001. *Move Reduces to Merge: A Theory of Locality*. Ph.D. thesis, University of Geneva. Available at <http://ling.auf.net/lingBuzz/000002>.
- Starke, Michal. 2011. Towards an elegant solution to language variation: Variation reduces to the size of lexically stored trees. Ms. University of Tromsø, available at ling.auf.net/lingBuzz/001183).
- Svenonius, Peter. 2004. On the edge. In *Peripheries: Syntactic Edges and their Effects*, edited by David Adger, Cécile de Cat, and George Tsoulas, pp. 261–287. Kluwer, Dordrecht.
- Svenonius, Peter. 2005. Extending the extension condition to discontinuous idioms. *Linguistic Variation Yearbook* 5: 227–263.
- Svenonius, Peter. 2007. Interpreting uninterpretable features. *Linguistic Analysis* 33 3-4: 375–413.
- Svenonius, Peter. 2010. The emergence of the functional sequence. Ms. CASTL, University of Tromsø; Paper presented on December 6, 2010 at the Christer Platzack Workshop, Lund University.
- Uriagereka, Juan. 1999. Multiple spell-out. In *Working Minimalism*, edited by Samuel David Epstein and Norbert Hornstein, pp. 251–282. MIT Press, Cambridge, Ma.
- Vijay-Shanker, K. and D.J. Weir. 1994. The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory* 27 6: 511–546.
- Wilder, Chris. 1999. Right-Node Raising and the LCA. In *Proceedings of the*, edited by Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest, pp. 586–598. Cascadilla Press, Somerville, Ma.
- Williams, Edwin. 2003. *Representation Theory*. MIT Press, Cambridge, Ma.