C O P *i* L CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **7**, Article 1: 1–31, 2015 | ISSN 2050-5949

RETHINKING FORMAL HIERARCHIES: A PROPOSED UNIFICATION*

THERESA BIBERAUER
University of Cambridge

IAN ROBERTS
University of Cambridge

ABSTRACT This paper addresses the question of the role and nature of formal hierarchies in current syntactic theory. Three types of hierarchy are highlighted: (i) cartographic-type hierarchies, (ii) formal feature hierarchies of the kind proposed by Harley & Ritter (2002), and (iii) parameter hierarchies of the kind first proposed by Baker (2001) and currently being elaborated on the Cambridge-based "Rethinking Comparative Syntax" project. We propose a unification of these three types, postulating a single formal hierarchy that is not predetermined by UG, but instead arises as an emergent property of the interaction of the three factors of language design introduced by Chomsky (2005), namely UG (F1), Primary Linguistic Data for language acquisition (PLD, F2) and third-factor considerations of cognitive computational conservativity (F3). We will also suggest some diagnostics for showing the unity of the hierarchies, with Relativised Minimality playing a central role in this connection.

1 Introduction

This chapter addresses the question of the role and nature of formal hierarchies in current syntactic theory. We identify three types of hierarchies: cartographic-type hierarchies of the kind that Cinque (1999) and Rizzi (1997) proposed for TP and CP respectively; featural hierarchies of the kind put forward by Harley & Ritter (2002); and parameter hierarchies of the kind first proposed

* The research reported here is funded by the European Research Council Advanced Grant No. 269752 "Rethinking Comparative Syntax". A shorter version of this paper is currently in press, Biberauer, T. & I. Roberts. Clausal hierarchies. To appear in: U. Shlonsky (ed.). *Beyond Functional Sequence.* Oxford: OUP. We thank Teresa Parodi, Ur Shlonsky, the native speakers who completed our English modals questionnaire, and the audience at the Comparative Syntax of English Workshop (Cambridge, November 2013) for their input. Usual disclaimers apply.

by Baker (2001) and currently elaborated by Roberts (2012), Biberauer & Roberts (2012a,b, 2014a,b,c), Biberauer, Holmberg, Roberts & Sheehan (2013), Biberauer (2013, 2014a,b), Sheehan (2013b) and other members of the ReCoS group (see Note 1 and www.mml.cam.ac.uk/dtal/research/recos).

As we will see, each of these hierarchies has its own empirical and conceptual strengths. But the obvious question arises as to how they are related. From a general perspective of methodological parsimony, it would be surprising if it turned out to be the case that there really were three distinct formal hierarchies, and the empirical case for this would have to be very strong. From a minimalist perspective in particular, we would not want, again unless the empirical case were irrefutable, to attribute such apparent complexity to UG or to the interacting factors contributing to the overall design of the language faculty.

Here we will suggest that these three hierarchies are in fact unified, and, moreover, that the single formal hierarchy is not predetermined by UG, but is an emergent property of the interaction of the three factors of language design introduced by Chomsky (2005), namely UG (F1), Primary Linguistic Data for language acquisition (PLD, F2) and third-factor considerations of cognitive computational conservativity (F3). We will also suggest some diagnostics for showing the unity of the hierarchies; here Relativised Minimality will play a central role.

2 Clausal hierarchies

As is well-known, Cinque (1999:90,106) presented an elaborate functional structure for the clause (i.e. the former IP/TP) of the following type:

(1) $\text{Mood}_{\text{Speech Act}}$ $\text{Mood}_{\text{Evaluative}}$ $\text{Mood}_{\text{Evidential}}$ $\text{Mod}_{\text{Epistemic}}$ T(Past) T(Future) $\text{Mood}_{\text{Irrealis}}$ $\text{Mod}_{\text{Necessity}}$ $\text{Mod}_{\text{Possibility}}$ $\text{Asp}_{\text{Habitual}}$ $\text{Asp}_{\text{Repetitive(I)}}$ $\text{Asp}_{\text{Frequentative(I)}}$ $\text{Asp}_{\text{Celerative(I)}}$ $\text{Mod}_{\text{Volitional}}$ $\text{Mod}_{\text{Obligation}}$ $\text{Mod}_{\text{Ability/Permission}}$ $\text{Asp}_{\text{Celerative(I)}}$ T(Anterior) $\text{Asp}_{\text{Terminative}}$ $\text{Asp}_{\text{Continuative}}$ $\text{Asp}_{\text{Perfect(?)}}$ $\text{Asp}_{\text{Retrospective}}$ $\text{Asp}_{\text{Proximative}}$ $\text{Asp}_{\text{Durative}}$ $\text{Asp}_{\text{Generic/progressive}}$ $\text{Asp}_{\text{Prospective}}$ $\text{Asp}_{\text{SgCompletive(I)}}$ $\text{Asp}_{\text{PlCompletive}}$ Voice $\text{Asp}_{\text{Celerative(II)}}$ $\text{Asp}_{\text{SgCompletive(II)}}$ $\text{Asp}_{\text{Repetitive(II)}}$ $\text{Asp}_{\text{Frequentative(II)}}$ $\text{Asp}_{\text{SgCompletive(II)}}$

Cinque's evidence for this hierarchy came from converging facts regarding the ordering of adverbs, auxiliaries and particles in many languages, and suffixes, particularly in agglutinating languages. More recently, Cinque (2006) has developed the hierarchy further, but the version given in (1) suffices for our purposes here.

Rizzi (1997) put forward an elaborated version of the left periphery (the earlier CP), splitting CP into ForceP, FocP, a possibly iterated TopP, and

FinP. The original proposal has been elaborated in various ways; Ledgeway (2010:51,(80)) summarises these developments with the following cartographic structure (see also i.a. Frascarelli & Hinterhölzl 1997, Speas & Tenny 2003, Giorgi 2010, Haegeman & Hill 2013, and much work by Halldór Sigurðsson):

(2) DeclP FrameP1 FrameP2 ConcP HypP ExclP ThemeP1 ThemeP2 IntP C-FocP1 C-FocP2 I-FocP1 I-FocP2 FinP

As Rizzi (1997) originally pointed out, it is not necessarily the case that all these categories are always active in every language (while, on the other hand, Cinque 1999:132-133 argues that this is the case for the hierarchy in (1); cf. also Cinque (2013)). We will develop Rizzi's observation in more detail below.

In contrast to the cartographic approaches, Chomsky (2000,2001) introduces the concept of "core functional categories", by which he means C, T and v in the clause. Chomsky (2005:18) speculates that "the more elaborate structures revealed by the cartographic inquiries are based on linearization of features in these [i.e. CP and vP, TB/IR] labels, and possibly labels closely linked to them (as in the C-T connection)". Here we propose to develop this idea by exploiting the distinction between formal and semantic features. The formal features are, as proposed in Chomsky (1995), interpretable or uninterpretable and, as such, are visible for syntactic operations such as Agree and Merge. The semantic features, on the other hand, are invisible to the core computational system (Narrow Syntax), but presumably visible at the semantic interface. We suggest that the formal feature that characterises the left periphery, for example, is C, which we take to be a generic clause-typing feature: hence the hierarchy in (2) is, for the Narrow Syntax, an iteration of C-positions. More precisely, this is a sequence of heads with the formal feature [C]. The features which distinguish the heads, and which the informal cartographic labels characterise, we take to be primarily semantic features, which may be grammaticalised to varying extents (see below). Similarly, the hierarchy in (1) can be seen as an iteration of T- and v-heads, with the v-domain starting at $\mathrm{Asp}_{\mathrm{Terminative}}$, the point of articulation between the "higher" and "lower" sequences identified by Cinque (1999:106).

Our proposal is more than just an attempt to reconcile the two seemingly disparate approaches to the structure of the clause. Instead, in the context of the parametric approach which we will discuss in §3, we see these approaches as reflecting two different levels of featural organisation. At a relatively low level of granularity, we can see the clause as consisting of (iterations of) Chomsky's core functional categories; at a higher level of granularity, we have the cartographic structures. Furthermore, these are not the only syntactically and semantically relevant levels of organisation in the clause. The phase level is higher than the core functional categories (since T, at least, is not inherently phasal); phases are clearly relevant to syntax and semantics, and, on many views, they are also computed in one way or another, at the PF interface (Marvin 2002, Newell 2008). A still higher level of organisation is the Extended Projection in the sense of Grimshaw (1991 *et seq.*, Shlonsky 2006). At this level, it may be that the only featural contrast is between verbs and nouns (i.e. either ±V or ±N); see also Biberauer, Holmberg & Roberts (2014).

Returning to the lower levels of organisation, it is common in cartographic work to refer to an informal notion of "field" or "zone". For example, Ledgeway (2010:41f.) refers to the topic and focus fields of the left periphery: contiguous groups of heads which share some semantic property and which can be ascribed a syntactic property also shared by other heads (C in this case).

To summarise, what we are suggesting is that clauses can be analysed at different levels of "magnification", as follows:[1]

(3) Extended Projection (V)> phase (C, v)> CFC (C, T, v) > "cartographic fields" (e.g. Tense, Mood, Aspect, Topic, Focus) > semantically/lexically distinct heads (as in (1) and (2)).

We conjecture that nominals have a similar organisation, but clearly disposing of different features, such as D, N, Num, Quantifier, etc. (cf. i.a. the papers in Cinque 2002, and Ihsane 2008 and Alexiadou, Haegeman & Stavrou 2009 in this connection) Adpositional phrases presumably also reflect this kind of organisation; cf. the papers in Cinque & Rizzi (2010) for discussion of the fine-grained cartographic structure of these categories.

Looking at (3), we observe that the features become ever more finely distinguished as we move down the hierarchy. It is widely held that the N- and V-features at the highest level, the Extended Projection, are active in the Narrow Syntax of all languages (see Chung 2012 and Biberauer 2014a for discussion). At the phasal level, C (and D, in the case of nominals) is cross-linguistically very common, but perhaps not universal (see Bošković 2008 on D), and clearly syntactically active when present (with corresponding absences of "activity" where it is absent – consider, for example, the difference between non-restructuring and restructuring infinitives in Germanic, where the former are clearly CPs). At the "core functional level", we have a broadly similar picture (we return to the point below). We will suggest that languages differ in the ways in which they grammaticalise the semantic features which define "cartographic fields". These features, we assume, are universally relevant to the semantic interface, but there appears to be considerable cross-linguistic variation as to how the lowest-level cartographic features are grammaticalised

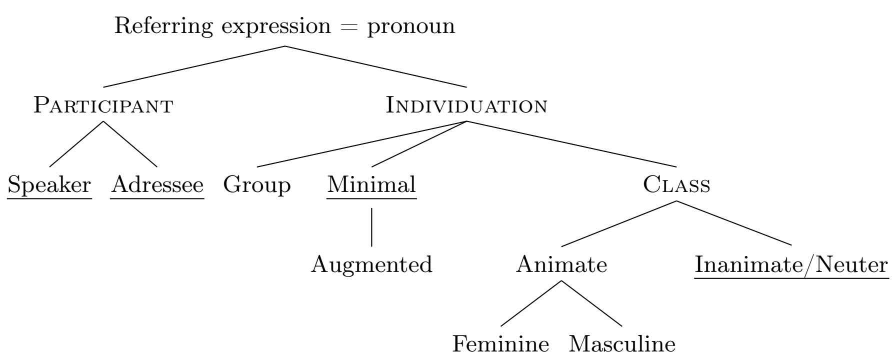
1 There may be a still lower level of "nanosyntactic" organisation (see Starke 2009). We will not elaborate in detail on how this level relates to the others just introduced, although we speculate briefly on this question below.

in different systems. It should be clear that these levels of organisation reflect a feature hierarchy. Ideally, then, we would like a non-trivial connection between the observations made here and the nature of syntactic feature hierarchies. We now turn to this question.

3 Feature hierarchies

In the recent literature, there are some examples of syntactic feature hierarchies, although none make any claim to be exhaustive, even in the context of the specific domain with which they are concerned. Harley & Ritter (2002:486), for example, present the following morphosyntactic feature hierarchy for pronouns (underlining indicates default features/values, which are assumed not to be formally represented in the Narrow Syntax in systems in which they are not contrastive, thus facilitating a node-counting metric for markedness – Harley & Ritter 2002:490):

(4)

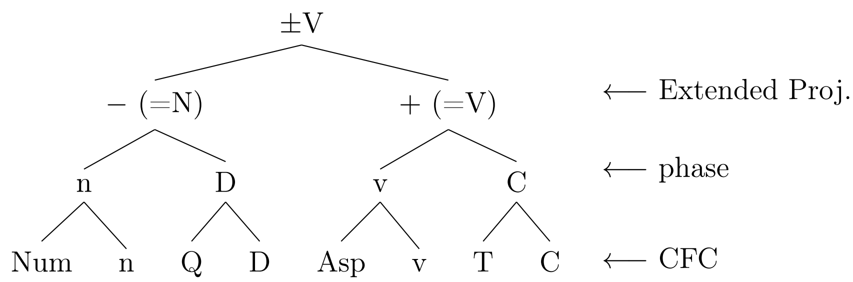


This hierarchy attempts to offer an explicit characterisation of the (hierarchical) dependencies between the [person], [number] and [gender] features standardly referred to under the umbrella-term, *ϕ-features*. As presented in Harley & Ritter (2002:487-488), the conceptual distinction between discourse-dependence and independence is what determines the structure of (4): the participant node dominates DP-properties anchored to discourse-roles, while the individuation node dominates DP-properties fixed independently of deictic considerations (e.g. [number] and [gender]).

Importantly, Harley & Ritter make some interesting and relevant claims about language acquisition in relation to the hierarchy in (4), namely: “UG provides a minimal initial structure which is elaborated in a deterministic fashion in response to contrasts detected in the input. Acquisition proceeds from the top down (i.e. from the root node): a given node must be acquired before its dependents” (498). Below we will elaborate on aspects of this approach to parameter-setting in acquisition.

The similarity between the categorial hierarchy we proposed in (3) and Harley & Ritter's (4) can be shown by reformulating (3) as follows:[2]

(5)



The hierarchy in (5) is evidently somewhat similar to that in (4). It is obviously incomplete, in that it must break up further into the more specified cartographic fields, and ultimately to the individual cartographic heads, but we leave that further (and practically more difficult) facet of (5) aside here. Our key point is that we can identify overall similarities between categorial and featural hierarchies of the type that have been proposed in the literature to date. In fact, if we take the categories in (5) to be features, then these may just be two different parts of the same hierarchy, with the pronominal hierarchy in (4) being distributed between the nP and DP phases.

Assuming that hierarchical structures such as those in (4) and (5) can represent the learning path in that the learner traverses the trees by "moving down" from the root node to lower nodes following branches of the tree, both (4) and (5) are, we contend, inadequate as representations of the learning path. This is so because these trees are both left- and right-recursive, creating multiple learning paths, implying, without further stipulation, that the trees are traversed in parallel, in a non-deterministic way. However, we would like to maintain that, at least at the earlier stages, acquisition is deterministic. As pointed out by Rice & Avery (1995:25) in connection with phonological acquisition, the early stages of acquisition show "global uniformity": this "refers to the observation that all children 'acquire roughly the same set of basic sounds in roughly the same order'" (Rice & Avery, quoted in Harley & Ritter 2002:498). Hence these trees present at least some components of the learning path in the wrong form. What is required from the acquisition perspective, then, is a hierarchy which at least partially features just one recursive branch, such that

2 Here we postulate Num, Q and Asp for concreteness, although it seems quite reasonable to extend the inventory of CFCs in this way, going beyond the specific proposals in Chomsky (2000,2001).

the relevant options represent deterministic endpoints to acquisition. We leave that task for future work, but illustrate below what part of a deterministic acquisition hierarchy could look like.

4 Parameter Hierarchies

As mentioned in the Introduction, the ReCoS project is developing the idea that parametric variation is an emergent property of the interaction of an underspecified UG, the PLD and third-factor computational conservativity on the part of the acquirer. For present purposes, the two principal linguistic manifestations of the acquirer's general computational conservativity are Feature Economy (FE) and Input Generalisation (IG) (the non-language-specific third factor at work here can be thought of as the general imperative to "make maximal use of minimal means" cf. Biberauer 2011, 2014a,b). These can be defined as follows:

(6) i. **Feature Economy** (FE) (generalised from Roberts & Roussou 2003:201):
Postulate as few formal features as possible to account for the input.

ii. **Input Generalisation** (IG) (adapted from Roberts 2007:275):
If a functional head F sets parameter P_j to value v_i then there is a preference for all functional heads to set P_j to value v_i.

From an acquirer's perspective, FE requires the postulation of the minimum number of formal features *consistent with the input.* IG embodies the logically invalid, but heuristically useful inference mechanism of learning from an existential to a universal generalisation. Like FE, it is stated as a preference, since it is always defeasible by the PLD. More precisely, we do not see the PLD as an undifferentiated mass, but we take the acquirer to be sensitive to particular aspects of PLD such as movement, agreement, etc., readily encountered in simple declaratives, questions and imperatives (see Biberauer 2014a,b and Note 4). So we see that the interaction of the second (PLD) and third-factor-derived (FE, IG) factors is crucial.

It may seem as though IG will create superset traps for the acquirer, but this is not the case if we think of the acquirer as overgeneralising due to ignorance of categorial distinctions (cf. Biberauer 2011, Branigan 2011,2012). This ignorance gradually erodes through the learning process, as finer and finer distinctions are made as a consequence of the interaction of all three factors: UG leaves certain options open (essentially many aspects of the formal-feature inventory, see below), the PLD provides evidence as to which options are needed, and FE and IG ensure that the maximally conservative options are always preferred, but that the formal distinctions required to capture the observed syntactic patterns are introduced during the acquisition process.[3]

In these terms, we can define parameter hierarchies. The hierarchy in (7) presents a highly simplified take on the parametric options relating to word order/linearization (see Biberauer 2013, 2014b for discussion).

(7) Is head-final present?
- No: **head-initial**
- Yes: present on all heads?
  - Yes: **head-final**
  - No: present on all [+V] heads
    - Yes: **head-final in the clause**
    - No: present on a subset of [+V] heads? . . .

Here "head-final" refers to the presence of a diacritic ˆ which, when associated with the categorial feature of a head, triggers movement of the complement of that head to its specifier, i.e. to what Biberauer, Holmberg & Roberts (2014) refer to as "L(inearisation)-movement".[4] This implies that head-final orders are formally marked in relation to head-initial ones.[5] The main evidence for this

3 Biberauer (2011, 2013, 2014a) and Biberauer & Roberts (2014a,c) highlight specific aspects of the PLD that will signal to acquirers that their current grammar lacks formal distinctions required to account for observed formal regularities in the target grammar, i.e. that what we might think of as a superset system needs to be refined. The driving intuition is that children are sensitive to systematic departures from Saussurian arbitrariness (arbitrary form-meaning mappings), and that the formal feature template given by their underspecified UG drives them to postulate formal features that capture these departures in a crosslinguistically parallel manner. Cues signaling departure from Saussurian arbitrariness include agreement and doubling more generally (children are known to be sensitive to morphology); movement (which assumes sensitivity to "basic" (neutral) word order, which, of course, also constitutes a departure from Saussurian arbitrariness; children are, again, known to be sensitive to this); systematic silence (e.g. null forms and ellipsis); and multifunctionality (superficial homophony); cf. also Zeijlstra (2008) and Wiltschko (2014) for further consideration of the formal significance of some of these properties.

4 In Biberauer, Holmberg & Roberts' terms, ˆ is a diacritic which associates with different formal features to trigger movement of different types: categorial feature-related ˆ triggers head-finality (L-movement); φ-feature-related ˆ triggers A-movement, and so on. See Biberauer, Holmberg & Roberts (2014) for more discussion.

5 The notion of formal markedness we adopt should not be taken to imply relative frequency; hence we are not predicting that head-initial languages are more frequent than head-final ones. Our notion of formal marking here relates simply to formal properties visible to the computational system, such as Case, for example; more generally, grammaticalised features are marked in the sense we are concerned with here.

stems from the Final-over-Final Constraint; see Biberauer, Holmberg & Roberts (2014). It should be readily apparent from (7) that the parameter hierarchy resembles the other hierarchies we have looked at in the previous sections, notably in that it makes reference to finer and finer categorial distinctions as we move downwards.

In fact, it is possible to isolate four main classes of parameter in terms of hierarchies of this type. Biberauer & Roberts (2012b) refer to these as **macro-, meso-, micro-**and **nanoparameters**. We can roughly characterise each class of parameter as follows:

(8) For a given value v_i of a parametrically variant feature F:

a. **Macroparameters**: all heads of the relevant type, e.g. all probes, all phase heads, etc, share v_i;

b. **Mesoparameters**: all heads of a given natural class, e.g. [+V] or a core functional category, share v_i;

c. **Microparameters**: a small, lexically definable subclass of functional heads (e.g. modal auxiliaries, subject clitics) shows v_i;

d. **Nanoparameters**: one or more individual lexical items is/are specified for v_i

Harmonically head-final order is therefore the consequence of a macroparameter. Japanese, Korean and the Dravidian languages are all well-known examples of harmonically head-final languages, and so they instantiate the macroparametric option at the top of the hierarchy in (7).[6] The null-subject parameter, as manifested for example in Latin and (most) Romance languages is a good example of a mesoparameter, since on standard formulations (e.g. Rizzi 1982, Holmberg 2005) it is determined by formal features of finite T and of pronominal Ds. Another likely case is Verb-Second in Germanic, which fairly clearly reflects properties of C, although the precise nature of these remains somewhat unclear.[7] The notion of microparameter defined in (8c) corresponds fairly closely to what is familiar in the recent literature on parameters; see in particular the discussion in Baker (2008). Finally, nanoparameters differ from the other kinds in that they represent synchronically systematic options that affect only a very small number of lexical items (possibly only one); they often correspond to diachronically earlier productive options. Crucially, nanoparameters must be distinguished from what one might think of as parametric fossils, i.e. syntactically inert lexicalisations which reflect a previously productive process,

6 Following the general view of parametric change as involving reanalysis of PLD through language acquisition (see Lightfoot 1979), we would expect such macroparameters to be diachronically very stable. This point is developed further in Biberauer & Roberts (2012b).

7 Verb-Second can be thought of as a mesoparameter since it crucially involves phasal C.

but no longer participate in syntactic operations. Nanoparametric settings tend to diachronically "regularise", or disappear, unless sufficiently frequent in the PLD. Many, perhaps all, nanoparametric options fall outside the core system defined by the parametric hierarchies under discussion here. To the extent that nanoparametric options involve high-frequency elements, they appear to be acquired as independent lexical items, independently of the more general properties of the system to which they belong; hence the much-discussed U-shaped acquisition pattern associated with the acquisition of high-frequency irregular verbs (cf. Marcus et al. 1992). In our terms, forms of this type would therefore not be acquired as a result of progessing down a given hierarchy, although their connection to specific hierarchies – in the sense that they appear to represent isolated instantiations within a given system of a pattern that can be seen to hold more systematically in other systems – is clear.

Parameter hierarchies clearly have the capacity to encode typological and diachronic generalisations. As such, they are, at the very least, useful descriptive and heuristic devices. However, taking seriously our emergentist view of parameters as described above, and bringing language acquisition into the picture, their explanatory potential becomes apparent. To see this, consider again the hierarchy in (7). The first choice-point apparently concerns fully harmonic head-initial vs head-final ordering. However, we do not see this as a UG-prescribed choice; instead, harmonic head-initial order represents a default option, since the L-movement trigger ^ is not present, in line with Feature Economy. Input Generalisation is also vacuously satisfied because (L-movement-related) ^ is entirely absent in the system. Most importantly, when the acquirer "takes the head-initial option" we do ***not*** consider this to be the result of the child interrogating the PLD with a view to assigning or not assigning ^ to some set of heads (including possibly the empty set). Instead, it is the result of the interaction of the three factors, which renders it unnecessary for the child to consider the relevance of L-movement-related ^ to this type of target grammar: UG, the first factor, clearly allows ^ to be absent since each head is underspecified in relation to this feature; the PLD (F2) will be compatible with the head-initial grammar and, further, both FE and IG are compatible with the head-initial grammar because no feature is posited and the absence of the feature is maximally general. Where the PLD contains no instances of head-finality, then, questions in relation to the hierarchy in (7) do not arise, with the result that head-initiality falls out as a consequence of the heads in the system not being specified in relation to a feature-type (the L-movement trigger) which is potentially available to grammatical systems. However, if the PLD (F2) is not compatible with the fully head-initial grammar (i.e. if the acquirer is forced to parse input strings as head-final), then the question of (L-movement-related) ^ (F1) arises, IG (F3) becomes relevant, and ^ is assumed to apply throughout the system so that a fully head-final system emerges (recall that the IG and FE are both defeasible by the PLD and here FE is clearly overridden). Importantly, no knowledge of categorial distinctions on the part of the acquirer is needed for convergence on either a fully head-initial or a fully head-final system.

At the next step "down the hierarchy", the connection with the categorial hierarchies discussed in §1 starts to be seen. If the PLD is such that generalised head-final order cannot be maintained (i.e. if it is such that some strings must be parsed as head-final and others as head-initial), then the simplest categorial distinction, namely [±V], is posited (with FE again overridden) and generalised (IG). A featural distinction of some kind arguably has to be in place if it is correct that Merge is driven by formal features (cf. i.a. Pesetsky & Torrego 2006, and see Biberauer 2014b on the more general matter of the crucial nature of some kind of basic categorial split). Nonetheless, as stated above, this distinction is not needed in the encoding of linearization properties unless the PLD is "mixed" in the way just described. Given "mixed" PLD, the acquirer "redeploys" this distinction in order to make its grammar compatible with the PLD; we can think of this as a kind of "categorial bootstrapping".

If there is no aspect of available PLD which is incompatible with the acquired grammar, then the acquisition process effectively halts and the acquirer converges on that system. Hence we expect to find languages which conform to each position on the hierarchy; in the present case, where [+V] Extended Projections show head-initial order and [–V] head-final order, may be instantiated for example by Thai (Simpson 2005) and Gungbe (Aboh 2004). The next level of the categorial hierarchy shown in (3) concerns the phase-heads C and v. If the PLD is such that generalised order (either head-final or head-initial) inside the verbal Extended Projection cannot be maintained (i.e. if it is such that some strings must be parsed as head-final and others as head-initial), then the next categorial distinction is posited. If the parametric and categorial hierarchies are unified, then this will be the distinction between C and v. We can think of this distinction, following Chomsky (2004:7), as the grammaticalisation of duality of semantics: vP is the domain of argument structure/ϑ-role assignment, and CP is the domain of discourse semantics (topic, focus, clause-type, etc). In acquisition terms, it is plausible to postulate that a formal distinction between the heads defining the argument and discourse domains is acquired early (i.e. that some kind of distinction between v and C arises) as acquisition studies suggest that children draw an early distinction between predicates and arguments[8] (which partially overlaps with and plausibly contributes to their initial [±V] distinction) and also clearly show that initial wh-placement is in place very early (see Valian & Casey 2003 for overview discussion). Returning to the formal characterisation of the clausal heads we are concerned with here: we can think that [+V] characterises the v-phase and [+V, +C] characterises the C-phase. So we see that the introduction of phase-heads involves adding just one feature, here [C].[9] Assuming clausal acquisition to proceed in "bottom-up" fashion, we understand why the "new" feature is added to the higher phase. Again, we expect that, to the extent that a system with different head-complement orders in the two clausal phases can be acquired, we should find languages that show different orders in CP and vP. Various West African languages (e.g. the Mande languages and the Senufo languages of the Gur group) appear to instantiate this possibility: they are clearly head-final in the vP, C is clearly initial and T is initial (cf. Creissels 2005). The inverse possibility, i.e. head-initial vP and head-final in CP is ruled out by the Final-over-Final Constraint (FOFC; see Biberauer, Holmberg & Roberts 2014; Sheehan 2013a). FOFC can be informally stated as follows:

(9) **The Final-over-Final Constraint (FOFC) (informal statement)**
A head-final phrase βP cannot dominate a head-initial phrase αP where α and β are heads in the same Extended Projection.

(9) rules out configurations of the following general kind:

(10) $*[_{\beta P} \ldots [_{\alpha P} \ldots \alpha \;\; \gamma P\,] \;\; \beta\hat{} \;\; (\alpha P)\,]$

In (10), α has no ˆ and hence its complement γP remains in complement position, while β has the ˆ-feature and so αP moves to the left of β, creating a specifier and giving rise to a configuration in which the head-initial αP is the complement of the head-final βP. In the clausal Extended Projection, α could be the lexical verb and β v. If auxiliaries can be merged in v, then (10) would instantiate the cross-linguistically unattested order V>O>Aux (see Biberauer in progress for discussion of apparent counterexamples; these turn out, on closer inspection, to obey (9)). Biberauer, Holmberg & Roberts propose that FOFC follows from the interaction of the antisymmetric nature of UG (i.e. the

8 Thanks to Teresa Parodi for discussion of this point, currently being developed in further research.

9 This oversimplifies since languages seem to distinguish between verbalising v and the various types of adicity-related v (see i.a. Harley 2012 for recent discussion). Taking this distinction into account, it is probably more correct to analyse verbalising v as the introducer of [+V], phasal v as the introducer of an argument-structure-related [v]-feature, and phasal C as the introducer of a clause-typing-related [C]-feature.

fact that head-final order requires ^[10]), combined with the highly local nature of c-selection, and Relativised Minimality (RM; see Rizzi 1990, 2001, 2013); see Biberauer, Holmberg & Roberts 2014, for a full discussion and explanation of how these assumptions derive the constraint. The latter two – highly local c-selection and RM – may be third-factor determined conditions on minimal search (see Ortega-Santos 2011), in which we again see the interaction of factors in language design,

The next level of the categorial hierarchy concerns the core functional categories (CFCs). If the PLD is such that generalised order (either head-final or head-initial) inside each clausal phase cannot be maintained (i.e. if it is such that some strings internal to each clausal phase must be parsed as head-final and others as head-initial), then the next categorial distinction is posited. If the parametric and categorial hierarchies are unified, then this will be the CFCs. As we saw, Chomsky (2000,2001) classifies C, T and v as the clausal CFCs. Hence, the further category which is introduced at this point is T. Following Ritter & Wiltschko (2009, in press) and Wiltschko (2014), we assume that tense, the ability to relate events to temporal intervals, is not necessarily grammaticalised as a structurally autonomous (component of a) functional head. [Tense] may not be grammaticalised at all – leaving another substantive (semantic) feature to be grammaticalised as the "anchoring" feature defining INFL (see Ritter & Wiltschko 2009, in press and Wiltschko 2014 for discussion) – or it may be grammaticalised, but located on the higher phase head (C), rather than defining a non-phasal CFC in the higher clausal domain (see Chomsky 2008 for a proposal along these lines, and Ouali 2008 on the idea that C may – parametrically – choose to KEEP features that could potentially be SHAREd or DONATEd to a non-phasal head within its domain). However, if the PLD provides evidence that a tense-marking element (e.g. verb or auxiliary) exhibits properties which require the postulating of a structurally distinct T-projection, [tense] is interpreted as a syntactically active formal feature which defines a projecting head (standardly labelled 'T'). Once T is postulated, it may, like any category, bear ^. Because of FOFC, however, it cannot bear ^ if v does not (see Biberauer 2013, 2014c on how this can also be shown to fall out without stipulation if one assumes the kind of three-factors model outlined here).

A question which naturally arises at this point is why the "extra" CFC shows up in the higher phase. We suggest that this is not necessarily the case. It is possible for a further CFC (not posited by Chomsky) to be posited in the vP phase, e.g. Asp, the locus of [aspect], i.e. grammaticalised aspect. This leads to the prediction that some languages grammaticalise tense and not

10 See Biberauer, Holmberg, Roberts & Sheehan (2010) and Biberauer, Roberts & Sheehan (2013) for a more articulated account of why phrase structure should be antisymmetric.

aspect and others have Asp and not T, while still others may grammaticalise both or neither. As far as we are aware, this is correct: a T-only language would be Afrikaans, where there is no simple grammatical expression of progressive or perfective aspect; Asp-only languages include Gungbe (Aboh 2004) and various Bantu languages (Nurse 2000:96), where there is rich aspect marking, but no clear autonomous tense-marking; languages grammaticalising both tense and aspect include English and Romance, while basilectal creole varieties may grammaticalise neither (if systems in which tense and aspect are exclusively lexicalised via adverbs are analysed as featuring only *semantic* and not *formal* tense and aspect features; cf. Biberauer 2011).

The next level of the categorial hierarchy concerns cartographic "fields" or regions. If the PLD is such that generalised order inside each CFC cannot be maintained (i.e. if it is such that some strings internal to each CFC must be parsed as head-final and others as head-initial), then the next type of categorial distinction is activated: cartographic fields. For example, within the C-system of various languages, certain particles are head-initial and others are head-final (cf. Davison 2007 on Indo-Aryan, Lee 2005 on San Lucas Quiavini Zapotec, and Willson 2005 on Marshallese); FOFC rules out a large subset of orders, namely those in which a head-final particle takes as its complement a head-initial particle. In principle, the same processes underlie the acquisition of finer-grained cartographic distinctions of the kind illustrated in (1), although we will not go into this here. In parametric terms, the cartographic level of granularity corresponds to the microparametric level of variation.

As noted above, nanoparametric variation, on the other hand, is in a sense outside the system in that it involves single lexical items whose idiosyncrasies simply have to be lexically encoded and whose properties are *not* generalised. IG therefore does not come into play, but FE presumably does.

Overall, then, we suggest that the parametric hierarchies approach facilitates a way of understanding the connection between some of the key formal hierarchies that have been proposed by generative syntacticians. Being grounded in the assumption that parametric distinctions generally and therefore categorial distinctions specifically are emergent, it furthermore provides insight into how we might productively think about the question of 'default' properties: as indicated above, at least some defaults can be seen to emerge where the interaction of Chomsky's three factors does not lead the acquirer to "ask" a question in respect of a higher-level formal property (i.e. one potentially defining various sub-types of feature; cf. also Biberauer, Roberts & Sheehan 2013 on the role of so-called 'no-choice' parameters in relation to defaults).

5 Relativized Minimality and the clausal hierarchy

In the previous sections, we have given a series of conceptual arguments for the view of the clausal hierarchy that we adopt. In this section, we will try to give empirical arguments for this. In all cases, the key mechanism is Relativized Minimality. What we will try to show is that intervention effects operate at different levels of granularity according to the levels of our hierarchy.

We adopt the following general formulation of Relativized Minimality (RM) from Rizzi (2001,2013):

(11) In the configuration
X ... Y ... Z ...
where each element asymmetrically c-commands the next, going from left to right, Y prevents X from interacting with Z for property P just where X and Y both have property P.

Following Starke (2001), we take P to refer to some (possibly composite) featural property (see also Rizzi 2013). Rizzi restricts RM to chain links; however, in keeping with the above speculation that it may derive from a third-factor constraint, we take it to be a general constraint on the interaction of formal features.

RM effects at the macro-level would, in relation to the kind of macro property we have been considering, have to involve X and Y in the configuration in (11) having the same value for [+/-V], thus creating massive intervention effects. Possible cases of this type include the traditional Head Movement Constraint (Travis 1984) and, in the "dominance" dimension (i.e. where X dominates Y in (11)), the A-over-A Condition. We will not go into these cases here, though.

RM effects are also observed at the meso-level. One important case concerns the "bottle-neck in C" which arguably gives rise to second-position effects. The idea, originally due to Haegeman (1996), and developed in Roberts (2004), is that movement of one category to a low specifier in the left periphery blocks movement of any other XP to any higher specifier position. Combined with movement of a verb or clitic to the relevant low position in the left periphery, this can give rise to a second-position effect (cf. Roberts 2011). Adopting for concreteness the schema for the left periphery in Rizzi (1997), the situation is as in (12):

(12) $[_{\text{ForceP}}\ [\ \cdots\ [_{\text{FinP}}$ XP V-/Cl-Fin $[_{\text{TP}} \ldots$ YP $\ldots$ (XP) $\ldots]]]]$

In our terms, a "bottle-neck effect" in (strict) V2 languages arises because the positions in the left periphery are not formally distinct from one another: the interaction of the PLD and FE leads the child to postulate an undifferentiated (root) C-head which triggers finite-verb movement and, via a feature-blind ˆ which targets the (outermost) vP-edge, movement of a single XP (cf. Chomsky 2008 on successive-cyclic movement). This is the simplest analysis of the Verb-Second input as it requires the acquirer to postulate only a single left-peripheral head which is not specifically associated with formal features like [topic] or [focus]. RM is relevant here as merging two (root) Cs in sequence would always, as a result of the higher C being featurally identical to the lower one, result in the initially moved XP blocking further phrasal movements into the CP-domain (the assumption here being that phase-heads bearing feature-blind ˆ always target the (outermost) edge of the next phase-head down; cf. again Chomsky 2008). As such, double-CP structures are predicted not to be postulated by Verb-Second-acquiring children: they will always be structurally redundant.[11] The situation is very different in non-V2 systems, where the C-system makes finer categorial distinctions, creating more positions and finer-grained RM interactions, so that there is no "bottle-neck" effect; see below for discussion of Italian and English in this connection. Strikingly, children do receive input signalling the need to postulate recursive merger structures of this type – consider (13), for example:

(13) a. We **all** could have been being punished for our crimes.
b. We could **all** have been being punished for our crimes.
c. We could have **all** been being punished for our crimes.
d. We could have been **all** being punished for our crimes.
e. * We could have been being **all** punished for our crimes.
(Harwood 2012:7)

Here the permitted stranding sites can be viewed as specifiers of formally identical heads, in this case, Ts (cf. Ramchand & Svenonius 2014 for independent arguments that the heads above that hosting progressive *-ing* – located somewhere below Cinque's $\mathrm{Asp}_{\mathrm{Perfect}}$ in (1) – should be viewed as heads associated with the TP-domain). In our terms, all these heads bear the formal

11 Strikingly, the Verb-Third (V3) structures that feature in Verb-Second languages (cf. Poletto 2010 and Holmberg 2013) are compatible with this analysis as they involve either elements that are plausibly first-merged at the edge of the CP-domain, meaning that no movement is required, or, in the case of Contrastive Left Dislocation, two elements that plausibly originated as a single "big DP", which can therefore be analysed as involving sub-extraction/stranding within the CP-domain. The Norwegian varieties considered by Westergaard (2009) pose a challenge to the view that Germanic-style V2-systems necessarily fail to distinguish formally between Cs of different kinds, suggesting that there is variation internal to the class of V2 languages, even Germanic-internally. There are various indications that North and West Germanic V2 may be more different than has typically been appreciated to date, but space considerations preclude consideration of this matter here.

specification [T], with [T], in turn, being associated with a formal-feature "bundle" including φ-features associated with ˆ. As such, all of the auxiliary positions above progressive *–ing* will be expected to attract the subject, accounting for the options in (13a-e); (13d), in turn, represents the position in which *all* is first-merged (cf. Harwood 2012 for a Late Merge analysis of floating quantifiers, in terms of which $\mathrm{v}_{\mathrm{prog}}$ is the v-phase-head, with passive *be* raising to this position); (13e) is then ruled out as it is below the First-Merge position for *all*. Here we conclude by highlighting the two diagnostics for mesoparametric effects that we have considered here (further diagnostics await future study): "bottle-neck effects" of the type observed in Verb-Second systems, and "recursive-merger" effects of the type seen in the English auxiliary system.

At the micro-level, finer-grained intervention effects are predicted to emerge. We can see this by contrasting movement to the left periphery in Italian and English, following Rizzi (2013). In Italian, multiple movements to the left periphery are allowed (see, for example, Rizzi 1997).

(14) A Gianni, QUESTO, domani, gli dovrete dire.
to Gianni.TOP this.FOC tomorrow.TOP you should tell.him
'To Gianni, THIS, tomorrow you should tell him' (Rizzi 1997:291)

In (14), we see that it is possible to move multiple topics and also a focused element to the left of the subject, a domain Rizzi (1997 *et seq.*) shows to be the CP-domain. The fact that these elements do not constitute interveners for one another, with the lower topic clearly not blocking movement of a further topic past it, shows clearly that the Italian left periphery contains more syntactically active heads than German does. Moreover, the fact that the ordering of the fronted elements is rigidly fixed suggests that a more fine-grained RM of the type initially proposed by Starke (2001) is in play here (see Rizzi 2013 for overview discussion and, i.a. Frascarelli & Hinterhölzl 2007 for discussion of how the two topic positions are formally distinct). In our terms, it is necessary not only to postulate multiple [C]-bearing heads to account for the data, but, more specifically, a sequence of [C]-heads bearing further formal features which must be merged in a fixed sequence (at least in part that dictated by semantic considerations; cf. also Ramchand & Svenonius 2014). Strikingly, multiple topicalisation is not possible in English:

(15) * John, your book, I convinced to buy.

This is because, in English, argument topicalisation necessarily involves a non-quantificational null operator, with the result that the double topicalisation in (15) is barred because the lower null operator blocks movement of the second null operator over it.Where one of the topicalised constituents is a modifier, however, multiple topicalisation is possible, as in (16):

(16) Words like that, in front of my mother, I would never say.

Here, the modifier *in front of my mother* does not contain a null operator and so does not block movement of the argument topic to the left periphery (see Rizzi 2013 for details and for further exemplification of interacting movements to the left periphery; on the English left periphery, see Haegeman 2012). Language-specific variation in relation to the formal features associated with CP-related heads and with discourse-marked XPs may therefore result in microparametric variation.

An interesting case of microparametric interaction involves the "modal field" in English. It is well known that English has a rich set of modal auxiliaries (see Palmer 1979, Huddleston & Pullum 2000, and Denison 1998 for a survey of changes in the recent history of the language). Palmer classifies the various uses of the core modals into epistemic, deontic and dynamic: consider the possibility, permission and ability senses of *can* respectively. Building on this classification, we characterise the modals as being lexically specified for the two formal features, which we will characterise as [±root] and [±∃].[12] Here, [+root] corresponds to the non-epistemic readings (ability/permission *can*, obligation *must*, volitional *will*, etc.), and [-root] corresponds to the various possibility and necessity readings. The feature [±∃], in turn, distinguishes "existential" (possibility/permission) from "universal" (necessity/obligation) readings, which, of course, are interdefinable by negation in the standard way. In these terms, *can* is, for example, lexically specified as [+∃,±root], *must* as [-∃,±root] and *might* as [+∃,-root], and so on (we return to the ± specifications below). Looking now at the subpart of the clausal hierarchy in (1) which relates to modality, given in (17), we expect there to be more modals than those which can be defined with reference to [±∃,±root]:

(17) $\text{Mood}_{\text{Speech Act}}$ $\text{Mood}_{\text{Evaluative}}$ $\text{Mood}_{\text{Evidential}}$ $\text{Mod}_{\text{Epistemic}}$ T(Past) T(Future) $\text{Mood}_{\text{Irrealis}}$ $\text{Mod}_{\text{Necessity}}$ $\text{Mod}_{\text{Possibility}}$ $\text{Asp}_{\text{Habitual}}\ldots$ $\text{Mod}_{\text{Volitional}}$ $\text{Mod}_{\text{Obligation}}$ $\text{Mod}_{\text{Ability/Permission}}$

By and large, what we see in practice is that the English modals can appear in any modal position that the two formal features just specified renders them semantically compatible with. The functional heads in the hierarchy can

12 [±root] and [±∃] should be understood as largely descriptive features, which we are employing as an interim measure, pending more detailed research on the extent to which these features can productively be viewed as instantiations of more basic features that may also be harnessed in other (non-modal) domains.

perhaps be thought of as specifying the modal base, in the sense of Kratzer (1981), with the extent to which these heads are *individually* present in different systems being determined by the extent to which the features that (uniquely) characterise them have been formalised (see below).

We can show how this works, and at the same time reveal some intricate micro- and nano-variation by following Cinque's method for establishing Merge-positions in the hierarchy. In terms of this method, we expect different interpretations of modals to relate to particular positions in the clause which are more readily identifiable in the presence of diagnostic adverbials. If we insert a given modal in the syntactic frame *Adv1 M Adv2* where each Adv is diagnostic of a potential head position in the hierarchy, then, we should obtain a picture of the extent to which the English modals occupy unique, dedicated positions (i.e. specialised formal-feature-bearing Mo(o)d heads) and the extent to which their distribution references less specific modal sub-fields (i.e. fewer Mo(o)d-heads). In what follows, we will consider structures of the type in (18), where HP and GP are contiguous Mo(o)d heads in the hierarchy in (17):

(18) $[_{\mathrm{HP}}$ Adv1 $[_{\mathrm{H'}}$ $[_{\mathrm{H}}$ M$]$ $[_{\mathrm{GP}}$ Adv2 ...$]]]$

Let us leave the two highest Mood heads ($\mathrm{Mood}_{\mathrm{SpeechAct}}$ and $\mathrm{Mood}_{\mathrm{Evaluative}}$) aside, taking them to be part of the C-system since they centrally involve speaker attitudes and hence are anchored to aspects of discourse semantics. In order to isolate the next-highest head, $\mathrm{Mood}_{\mathrm{Evidential}}$, we place our modals in between the adverbs *allegedly*, diagnostic of $\mathrm{Mood}_{\mathrm{Evidential}}$P, and *probably*, diagnostic of $\mathrm{Mood}_{\mathrm{Epistemic}}$P. Restricting ourselves to the "core" modals, i.e. excluding *shall, dare, need* and *ought* (all of which are in different ways peripheral to the contemporary English modal system; see Huddleston & Pullum 2000), and treating the historically past forms as distinct modals at least for illustration, the range of modals in question is *might, may, can, could, should, must, will* and *would*. In (19), we show each of these modals in the adverbial frame which isolates the putative $\mathrm{Mood}_{\mathrm{Evidential}}$ head, in a context which naturally favours an evidential interpretation:

(19) According to our latest readings, the Higgs Boson allegedly might/ $^{?B}$may/could/should/$^{?R,*B}$must/$^{??R,*B}$can/will/$^{?B,R}$would probably exist.

(Superscript "R" here indicates Roberts' judgement, and superscript "B" Biberauer's where a structure is not fully acceptable. There are intricate and intriguing patterns of regional and, probably, age-related variation at work here, mainly concerning the single item *may*, that will form the subject of a more detailed future investigation).

If we negate (19), we observe a striking difference in acceptability between contracted and non-contracted negation:

(20) According to our latest readings, the Higgs Boson ...

a. ... allegedly couldn't / $^{??\mathrm{R},*\mathrm{B}}$mightn't / $^{*\mathrm{R},\mathrm{B}}$mayn't / $^{?\mathrm{R},*\mathrm{B}}$shouldn't / $^{*\mathrm{R},\mathrm{B}}$mustn't /$^{??\mathrm{R},*\mathrm{B}}$can't/$^{??\mathrm{R},*\mathrm{B}}$won't/$^{*\mathrm{R},\mathrm{B}}$wouldn't probably exist.

b. ... allegedly might/$^{?\mathrm{B}}$may/could/should/$^{??\mathrm{R},*\mathrm{B}}$must/$^{??\mathrm{R},*\mathrm{B}}$can/will /$^{?\mathrm{R},?\mathrm{B}}$would probably not exist.

With all the modals except, for Roberts, *could*, contracted negation is impossible in this context. But, except for *mayn't* (which Palmer 1979 claims not to exist at all in contemporary English, although Denison 1998 takes issue with this; either way, it is not part of the English of either of the current authors), the forms with contracted negation do exist, as we shall see below. We can account for the contrast in (20) as follows: *not* is generally agreed to be a "low" negation element (see Tubau 2008 for recent discussion), merged within the low vP-domain, just above VP, while the contracted forms have convincingly been shown to be independent lexical items (cf. Zwicky & Pullum 1983) which seem good candidates for Merger at some stage of the derivation in the left-peripheral polarity-related phrase that i.a. Duffield (2007, 2011), Biberauer (2012), Kandybowicz (2013) and Batllori & Hernanz (2013) all argue on independent grounds to be located at the outermost edge of vP (Duffield 2011 specifically argues for a low modality-related position, which may help us to locate this XP in relation to (1)). We return to the significance of these data below.

A similar positive-vs-contracted modal discrepancy arises in relation to *might/mightn't* and *must/mustn't* (we leave *may/mayn't* aside, given the complications surrounding their use) if we isolate the modals in relation to a putative $\mathrm{Mod}_{\mathrm{Epistemic}}$ head. Consider (21):

(21)
a. This site probably might/$^{??\mathrm{B}}$may/could/$^{??\mathrm{R},*\mathrm{B}}$can/must/will/would once have been a Roman camp.

b. This site probably $^{*\mathrm{R},\mathrm{B}}$mightn't/$^{\mathbf{*R/B}}$mayn't/couldn't/$^{?\mathrm{R}}$can't/ $^{*\mathrm{R},\mathrm{B}}$mustn't/won't/wouldn't once have been a Roman camp.

c. This site probably might/$^{??\mathrm{B}}$may/could/$^{??\mathrm{R},*\mathrm{B}}$can/$^{??\mathrm{R}}$must/will /$^{?\mathrm{R}}$would once not have been a Roman camp.

Here the modals are placed between *probably*, associated with $\mathrm{Mod}_{\mathrm{Epistemic}}$ and *once*, associated with $\mathrm{T}_{\mathrm{Past}}$. Strikingly, the judgements in (21a) indicate that *might, could* and, for Biberauer, *would* are as compatible with merger in the

$\text{Mod}_{\text{Epistemic}}$-domain as they are with merger in the $\text{Mood}_{\text{Evidential}}$-domain (cf. (20a)).

Next, we place the modals in the T_{Past}-domain, between *once*, marking its left edge, and *perhaps*, marking the left edge of $\text{Mod}_{\text{Irrealis}}$-domain (we cannot test with an adverb such as *then*, probing T_{Future}, owing to semantic incompatibility with *once*):

(22) John once could(n't)/$^{*R,?B}$might/^{*R,B}may/^{*R,B}can/^{*R,B}must/^{*R,B}should perhaps speak French.

Perhaps unsurprisingly, the only modals able to appear here are ability *could* and *couldn't*, arguably the only past-form modal in contemporary English which is still able to function as the past tense of ability *can* (note that (22) shows that this is not true of *might* and *should*; these items are no longer the past forms of *may* and *shall*). The fact that the "genuine pastness" of *could* correlates with a placement option not available to the other historically past modals plausibly signals the presence of a formal [Past]/[Tense:Past] feature on *could*. Given the absence of this formal feature on the other modals under consideration, we can also understand the pattern in (22).

Turning to T_{Future}, we place the modals between *then* and *perhaps*:

(23) John then will/won't/would/$^{?B}$wouldn't/$^{?R}$might/$^{?R,B}$may/must/ $^{?R}$could/ $^{?R,B}$can/should/$^{?B}$shouldn't/^{*R,B}mustn't perhaps give up.

Here, most modals are to some extent grammatical (on a non-root interpretation), with *can* marginal and *mustn't* impossible. Given that *will* is usually seen as the "future auxiliary", and the future-orientation of non-root modality generally, this is unsurprising. That *mustn't* should be ungrammatical suggests that the non-root form of this lexical item has, like the root form, become featurally very specialised in ways that would take us too far afield here; like other more specialised forms, though, we see that this specialisation strongly restricts the number of modal domains in which *mustn't* may occur (cf. the discussion of the contracted forms above).

Next, consider $\text{Mod}_{\text{Irrealis}}$, located between *perhaps* and *necessarily*:

(24) John perhaps$^{?R,*B}$might/$^{?R,*B}$will/^{*R}must/$^{?B}$may/$^{?R}$should/$^{?R}$could /$^{??R}$can/would/won't necessarily finish his thesis, if he had/has the energy.

Here it appears that only *must* is incompatible with this position (although *can* is again quite marginal, and *might* is also dubious). This once again points to this modal being featurally specialised, significantly more so than most of the rest of the English modal inventory; we leave the details to future research.

Strikingly, the contracted negation forms are possible here, which may indicate that they encode [$\text{Mod}_{\text{Irrealis}}$] in addition to the low modal features already discussed.

Finally, let us consider $\text{Mod}_{\text{Necessity}}$ and $\text{Mod}_{\text{Possibility}}$:

(25) a. The children necessarily must/will/won't/ mustn't/^{*R,B}might/ ^{*R,B}may/^{*R,B}could/couldn't/$^{?R,*B}$mightn't/should(n't) usually go to bed early.

b. The children possibly must/mustn't/will/might/may/can/could/ should/won't/ $^{?R}$mustn't/$^{?R}$mightn't/shouldn't usually go to bed early.

Unsurprisingly, the positive forms of the "existential" modals *might*, *may* and *could* are not allowed in the $\text{Mod}_{\text{Necessity}}$-domain; the negative ones are, though, presumably owing to the logical equivalence of ¬POSS and NEC¬. As shown in (25b), all the modals, existential and universal, positive and negative, are allowed in the $\text{Mod}_{\text{Possibility}}$-domain, which suggests that this head really encodes a weak form of modality (if possibility is to be construed as existential quantification over possible worlds, then this is a case of weak quantification, unlike $\text{Mod}_{\text{Necessity}}$, which involves strong quantification).

In (26), we summarise the rather intricate observations of (19-25), indicating for each putative modal head which modals may realise the meanings associated with it (for these purposes, we treat "?" as compatible with the meaning in question):

(26) a. $\text{Mood}_{\text{Evidential}}$: *might, may, could, should* (non-root only)

b. $\text{Mod}_{\text{Epistemic}}$: *might, may, could, couldn't, must* (non-root only)

c. T_{Past}: *could, couldn't* (root only, i.e. ability)

d. T_{Future}: *will, won't, would, wouldn't, must* (non-root), *shouldn't* (non-root), *may*

e. $\text{Mod}_{\text{Necessity}}$: *must, will, won't, mustn't, couldn't, should(n't)* (root only)

f. $\text{Mood}_{\text{Irrealis}}$: *might, will, may, should, could, can, would, won't*

g. $\text{Mod}_{\text{Possibility}}$: *must, will, might, may, can, could, couldn't, should, won't, shouldn't*

The picture that emerges is that each modal appears in the functional heads which are compatible with its intrinsic lexical properties, specified in terms of the features [±root] and [±∃], with additional features presumably coming into play in certain cases, as indicated above. The precise interpretation of each (underspecified) modal appears to be the result of the interaction of these formal features and the modal base provided by the functional heads in the system. We have not specified the identity of these heads here (although Cinque's identification of the heads in the modal field has proven heuristically useful), but what the approach outlined in §3 leads us to conclude on the basis of the data considered in this section is that English is likely to have a featurally rather fine-grained modal system, i.e. what one would expect in parametric-hierarchy terms for a complex, idiosyncratic component of grammar, which is necessarily defined on the basis of quite specific featural distinctions. More specifically, what the data suggest is, on the one hand, that children will analyse the majority of the lexical items instantiating the English verbal modal system as underspecified elements: this will allow them to minimise the number of (verbal) modal lexical items they ascribe distinct lexical entries to, as featurally underspecified elements are expected to be compatible with a range of Merge positions; on the other hand, the various cases where we see a modal use which can independently be shown to exist being *unavailable* in a semantically compatible context suggest the presence of further formal features – features which must then be housed on functional heads, given our previous observation regarding the number of distinct lexical entries a three-factors-driven acquirer is likely to want to postulate. The behaviour of the contracted forms illustrated in (20) is a case in point: here we saw that modals compatible with $\mathrm{T}_{\mathrm{Future}}$ interpretations cannot surface in $\mathrm{Mod}_{\mathrm{Epistemic}}$, while those compatible with $\mathrm{T}_{\mathrm{Past}}$ interpretations (associated with the head Cinque 1999 assumes to be immediately subjacent to $\mathrm{Mod}_{\mathrm{Epistemic}}$) can. In RM terms, this gap can be understood as signalling the presence of a functional head whose featural specification is such that it is not possible for $\mathrm{T}_{\mathrm{Future}}$-compatible modals to raise across it. That $\mathrm{T}_{\mathrm{Future}}$-compatible modals can front into the C-domain in interrogatives (consider: *Wouldn't that be a good idea?*) also has the hallmark of RM: in this case, we would expect *wouldn't* to be associated with an additional feature by virtue of its role in the interrogative; in line with the refined take on RM initiated in Starke (2001), we can understand this as the consequence of the additional feature serving to render *wouldn't* featurally more specific than the intervening head, thus facilitating movement across a head that would otherwise intervene.

Clearly there is much, much more to say before we arrive at anything like a complete account of the English modals. However, that is not our main goal here; our goal was simply to show, firstly, that there evidently are empirical phenomena which cannot be understood without invoking fine-grained cartography, even in the context of the kind of emergentist approach to features we are advocating here, and, secondly, that RM appears to be fundamental to the diagnosis of formal structure at this fine-grained level, just as it is at higher levels.

6 Conclusion

In the foregoing, we have tried to indicate how three apparently distinct hierarchies, the clausal hierarchy, feature hierarchies and parameter hierarchies, may in fact be reduced to just one. This also offers a way to unify cartographic analysis with Chomsky's (2001) Core Functional Categories and Grimshaw's (1991) Extended Projections. Two key diagnostics reveal the nature of the hierarchies: types of parametric variation and Relativised Minimality. Both of these apply at differing levels of granularity in different languages. If we are right, the possible levels of granularity are the same in both cases.

In the foregoing, we have done little more than instigate a programme for further research. However, we consider this to be a very promising programme, one which may reveal a deep unity at the heart of syntax.

References

Aboh, Enoch. 2004. *The morphosyntax of complement-head sequences. Clause structure and word order patterns in Kwa.* Oxford: Oxford University Press.

Alexiadou, Artemis, Liliane Haegeman & Melita Stavrou. 2009. *The noun phrase in generative perspective.* Berlin: Mouton de Gruyter.

Baker, Mark. 2001. *The atoms of language: The mind's hidden rules of grammar*. London: Basic Books.

Baker, Mark. 2008. The macroparameter in a microparametric world. In: Theresa Biberauer (ed.). *The limits of syntactic variation.* Amsterdam: Benjamins, 351-74.

Batllori, Montserrat & Luisa Hernanz (2013). Emphatic polarity particles in Spanish and Catalan. *Lingua* 128: 9-30.

Biberauer, Theresa. 2011. In defence of lexico-centric parametric variation: two 3rd factor-constrained case studies. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).

Biberauer, Theresa. 2012. Predicate doubling in (Germanic) contact varieties. Introducing the case of Afrikaans. Paper presented CGSW 27 (Yale).

Biberauer, Theresa. 2013. Offers that cannot be refused: mafioso parameters and the construction of minimalist parameters. Paper presented at the

39th Incontro di grammatica generative (Modena).

Biberauer, Theresa. 2014a. The universal component of emergent categories. Paper presented at the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) Annual Conference (Marburg).

Biberauer, Theresa. 2014b. Pro-drop and emergent parameter hierarchies. Paper presented at the Understanding Pro-drop: A Synchronic and Diachronic Perspective Conference (Trento).

Biberauer, Theresa. 2014c. Particles, peripherality and the Final-over-Final Constraint. Paper presented at the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) Annual Conference (Marburg)

Biberauer, Theresa. In progress. FOFC and verb clusters. To appear in: Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan. *The Final-over-Final Constraint*. Cambridge, MA: MIT Press.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45(2):169-225.

Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan. 2010. Reconciling formalism and functionalism: a minimalist perspective. Presented at the LAGB Annual Meeting (Leeds).

Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan. 2013. Complexity in comparative syntax: the view from modern parametric theory. In: Fritz Newmeyer & Laurel Preston (eds). *Measuring linguistic complexity*. Oxford: Oxford University Press, 103-127.

Biberauer, Theresa and Ian Roberts. 2012a. Towards a parameter hierarchy for auxiliaries: diachronic considerations. In: J. Chancharu, X. Hu & M. Mitrović (eds). *Cambridge Occasional Papers in Linguistics* 6. 209-36.

Biberauer, Theresa and Ian Roberts. 2012b. On the significance of what hasn't happened. Paper presented at DiGS 14 (Lisbon).

Biberauer, Theresa and Ian Roberts. 2014a. Parameter setting. To appear in: A. Ledgeway & I. Roberts (eds). *The Cambridge Handbook of Historical Syntax*. Cambridge: Cambridge University Press.

Biberauer, Theresa and Ian Roberts. 2014b. Conditional Inversion and types of parametric change. To appear in a festschrift.

Biberauer, Theresa and Ian Roberts. 2014c. Contrastive Hierarchies in Phonology and Syntax: The Role of Third Factors (or: Why Phonology is *not* Different). Unpublished ms: University of Cambridge.

Biberauer, Theresa, Ian Roberts and Michelle Sheehan. 2013. No-choice Parameters and the Limits of Syntactic Variation. In: R. Santana-LaBarge (ed). *Proceedings of the 31st West Coast Conference on Formal Linguistics (WCCFL 31).* Somerville, MA: Cascadilla Press, 46-55.

Bošković, Željko. 2008. What will you have, DP or NP? *Proceedings of the North East Linguistic Society 37.* GLSA: University of Massachusetts, 101-114.

Branigan, Phil. 2011. In defense of macroparameters: Algonquian evidence. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).

Branigan, Phil. 2012. Macroparameter learnability: an Algonquian Case Study. Unpublished ms: Memorial University of Newfoundland.

Chomsky, Noam.1995. *The minimalist program.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In: Roger Martin, David Michaels & Juan Uriagereka (eds). *Step by step: essays on minimalist syntax in honor of Howard Lasnik.* Cambridge, MA: MIT Press, 89-156.

Chomsky, Noam. 2001. Derivation by phase. In: Michael Kenstowicz (ed.). *Ken Hale: a life in language.* Cambridge, MA: MIT Press, 1-53.

Chomsky, Noam. 2004. Beyond Explanatory Adequacy. In: Adriana Belletti (ed.). *Structures and beyond. The cartography of syntactic structures* (Volume 3). Oxford: Oxford University Press, 104-131.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1-22.

Chomsky, Noam. 2008. On Phases. In: Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds). *Foundational issues in linguistic theory.* Cambridge, MA: MIT Press, 133-166.

Chung, Sandra. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38: 1-56.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a crosslinguistic perspective.* Oxford: Oxford University Press.

Cinque, Guglielmo. 2002. *Functional structure in DP and IP. The cartography of syntactic structures* (Volume 1). New York: Oxford University Press.

Cinque, Guglielmo. 2006. *Restructuring and functional heads. The cartography of syntactic structures* (Volume 4). New York: Oxford University Press.

Cinque, Guglielmo. 2013. Cognition, typological generalizations, and Universal Grammar. *Lingua* 130: 50-65.

Cinque, Guglielmo & Luigi Rizzi. 2010. *Mapping spatial PPs. The cartography of syntactic structures.* Volume 6. Oxford: Oxford University Press.

Creissels, Denis. 2005. S-O-V-X constituent order and constituent order alternations in West African languages. *Berkeley Linguistics Society* 31: 37-51.

Davison, Alice. 2007. Word order, parameters and the Extended COMP projection. In: Josef Bayer, Tanmoy Bhattacharya & M.T. Hany Babu (eds). *Linguistic theory and South Asian languages.* Amsterdam: Benjamins, 175-198.

Denison, David (1998). Syntax. In: Suzanne Romaine (ed.). *The Cambridge History of the English Language*: *1776-1997* (Volume 4). Cambridge: Cambridge University Press, 92-329.

Duffield, Nigel (2007). Aspects of Vietnamese Clausal Structure: Separating Tense from Assertion. *Linguistics* 45: 765-814.

Duffield, Nigel (2011). Polarity Emphasis and 'Low Modality' in Vietnamese and English. Paper presented at GIST 4 (Ghent).

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In: Susanne Winkler & Kerstin Schwabe (eds). *On information structure, meaning and form.* Amsterdam: Benjamins, 87-116.

Giorgi, Alessandra. 2010. *About the speaker. Towards a syntax of indexicality.* Oxford: Oxford University Press.

Grimshaw, Jane. 1991. *Extended projection.* Unpublished ms: Rutgers.

Haegeman, Liliane. 1996. Verb Second, the Split CP and Null Subjects in Early Dutch Finite Clauses. *Geneva Generative Papers/GenGenP* 4(2). 133-175.

Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena and the composition of the left periphery. The cartography of syntactic structures* (volume 8). Oxford: Oxford University Press.

Haegeman, Liliane, and Virginia Hill. 2013. The syntactization of discourse. In: Raffaella Folli, Christina Sevdali & Robert Truswell (eds). *Syntax and its limits.* Oxford: Oxford University Press, 370-390.

Harley, Heidi. 2012. External arguments and the Mirror Principle: on the distinctness of Voice and v. *Lingua* 125: 34-57.

Harley, Heidi & Elizabeth Ritter. 2002. A feature-geometric analysis of person and number. *Language* 78(3). 482-526.

Harwood, Will. 2012. There are several positions available: English intermediate subject positions. In: Enrico Boone, Kathrin Linke & Maartje Schulpen (eds). *Proceedings of ConSOLE XIX:* 1-25. Leiden: LUCL.

Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36(4). 533-564.

Holmberg, Anders. 2013. Verb Second. To appear in: Tibor Kiss & Artemis Alexiadou (eds). *Syntax. An international handbook of contemporary syntactic research.* Berlin: Mouton.

Huddleston, Rodney & Geoffrey Pullum (2000). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Ihsane, Tabea. 2008. *The layered DP: form and meaning of French indefinites.* Amsterdam: Benjamins.

Kandybowicz, Jason (2013). Ways of emphatic scope-taking: From emphatic assertion in Nupe to the grammar of emphasis. *Lingua* 128: 51-71.

Kratzer, A. 1981. The notional category of modality. In: H. Eikmeyer & H. Rieser (eds). *Words, Worlds and Contexts.* Berlin: de Gruyter, 38-74. Reprinted in P. Portner & B. Partee (eds). 2002. *Formal semantics: The essential readings.* Oxford: Blackwell, 289-323.

Ledgeway, Adam. 2010. The clausal domain: CP structure and the left periphery. In: Roberta D'Alessandro, Adam Ledgeway & Ian Roberts (eds). *Syntactic variation. The dialects of Italy.* Cambridge: Cambridge University Press, 38-51.

Lee, Felicia. 2005. Clause-fronting and clause typing in San Lucas Quiaviní Zapotec. In: Andrew Carnie, Heidi Harley & Sheila Ann Dooley (eds). *On the Syntax of verb-initial languages.* Amsterdam: Benjamins, 91-106.

Lightfoot, David. 1979. *Principles of Diachronic Syntax.* Cambridge: Cambridge University Press.

Marcus, Gary, Steven Pinker, Michael Ullman, Michelle Hollander, John Rosen and Fei Xu. 1992. *Overregularization in language acquisition.* Chicago: University of Chicago Press.

Marvin, Tatiana. 2002. *Topics in the stress and syntax of words.* Ph.D. dissertation: MIT.

Newell, Heather. 2008. *Aspects of the phonology and morphology of phases.* Ph.D. dissertation McGill University.

Nurse, Derek. 2000. *African languages: an introduction.* Cambridge: Cambridge University Press.

Ortega-Santos, Ivan. 2011. On Relativized Minimality, memory and cue-based parsing. *IBERIA* 3(1). 35-64.

Ouali, Hamid. 2008. On C-to-T Phi-feature Transfer: the nature of agreement and anti-agreement in Berber. In: Roberta D'Alessandro, Gunnar Hrafn Hrafnbjargarson & Susann Fischer (eds). *Agreement Restrictions.* Berlin: Mouton de Gruyter, 159-180.

Palmer, Frank (1979). *Modality and the English Modals.* London: Longman.

Pesetsky, David & Esther Torrego. 2006. Probes, Goals and syntactic categories. In: Y. Otsu (ed.) *Proceedings of the 7th Annual Tokyo Conference on Psycholinguistics.* Tokyo: Hituzi Syobo Publishing Company.

Poletto, Cecilia. 2010. *The higher functional field. Evidence from northern Italian dialects.* Oxford: Oxford University Press.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the Functional Hierarchy. To appear in *Language Sciences.*

Rice, Keren & Peter Avery. 1995. Variability in a deterministic model of language acquisition. In: J. Archibald (ed.). *Phonological acquisition and phonological theory.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 23-42.

Ritter, Elizabeth & Martina Wiltschko (2009). Varieties of INFL: TENSE, LOCATION and PERSON. In: Jeroen van Craenenbroeck (ed.). *Alternatives to Cartography.* Berlin: Mouton de Gruyter, 153-201.

Ritter, Elizabeth & Martina Wiltschko (in press). The composition of INFL. An exploration of tense, tenseless languages, and tenseless constructions. To appear in *Natural Language and Linguistic Theory.*

Rizzi, Luigi. 1982. *Issues in Italian syntax.* Dordrecht: Foris.

Rizzi, Luigi.1990. *Relativized Minimality.* Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In: Liliane Haegeman (ed.). *Elements of grammar.* Dordrecht: Kluwer, 281-337.

Rizzi, Luigi. 2001. Relativized Minimality Effects. In: Mark Baltin & Chris Collins (eds). *The handbook of contemporary syntactic theory.* Oxford: Blackwell, 89-110.

Rizzi, Luigi. 2013. Locality. *Lingua* 130. 169-186.

Roberts, Ian. 2004. The C-System in Brythonic Celtic Languages, V2, and the EPP. In L. Rizzi (ed) *The Structure of IP and CP. The Cartography of Syntactic Structures.* Volume 2. Oxford: Oxford University Press, 297-328.

Roberts, Ian. 2007. *Diachronic syntax.* Oxford: Oxford University Press.

Roberts, Ian. 2011. Phases, head movement and second-position effects. In: Ángel Gallego (ed.). *Phases. Developing the framework.* Berlin: Mouton, 385-440.

Roberts, Ian. 2012. Macroparameters and Minimalism: A Programme for Comparative Research. In: Charlotte Galves, Sonia Cyrino, Ruth Lopes, Filomena Sândalo & Juanito Avelar (eds). *Parameter theory and linguistic change.* Oxford: Oxford University Press, 319-334.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change. A minimalist approach to grammaticalization.* Cambridge: Cambridge University Press.

Sheehan, Michelle. 2013a. Some implications of a copy theory of labeling. *Syntax* 6(4). 362-396.

Sheehan, Michelle. 2013b.Towards a parameter hierarchy for alignment. In: Robert Santana-LaBarge (ed.). *Proceedings of WCCFL 31.* Somerville, MA: Cascadilla Press, 399-408.

Shlonsky, Ur. 2006. Projection étendue et cartographie de SC. *Nouveaux cahiers de linguistique française* 27: 83-93.

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. In: Guglielmo Cinque & Richard Kayne (eds). *The Oxford handbook of comparative syntax.* Oxford: Oxford University Press, 806-838.

Speas, Margaret. 2004. Evidentiality, logophoricity, and the syntactic representation of pragmatic features. *Lingua* 114: 255-276.

Speas, Margaret & Carol Tenny. 2003. Configurational properties of Point of View roles. In: Anna-Maria di Sciullo (ed.). *Asymmetry in Grammar.* Amsterdam: John Benjamins, 315-344.

Starke, Michal. 2001. *Move dissolves into Merge: a theory of locality.* Ph.D. dissertation: University of Geneva.

Starke, Michal. 2009. *Nanosyntax. A short primer to a new approach to syntax.* Unpublished ms: Tromsø http://ling.auf.net/lingbuzz/001230, 25 April 2013

Travis, Lisa. 1984. *Parameters and effects of word order variation.* Ph.D. dissertation: MIT.

Tubau, Susagna. 2008. *Negative Concord and Romance: syntax-morphology* *interface conditions on the expression of negation.* Ph.D. dissertation: Utrecht.

Valian, Virginia & Lyman Casey. 2003. Young children's acquisition of *wh*-questions: the role of structured input. *Journal of Child Language* 30. 117-143.

Westergaard, Marit. 2009. *The acquisition of word order. Micro-cues, information structure, and economy.* Amsterdam: John Benjamins.

Willson, Heather. 2005. Marshallese *yes/no* questions and remnant movement. *UCLA Working Papers in Linguistics: Proceedings of the 12th Meeting of the Austronesian Formal Linguistics Association* 12: 421-435.

Wiltschko, Martina. 2014. *The universal structure of categories. Towards a formal typology.* Cambridge: Cambridge University Press.

Zeijlstra, Hedde. 2008. On the syntactic flexibility of formal features. In: Theresa Biberauer (ed.). *The limits of syntactic variation.* Amsterdam: John Benjamins, 143-173.

Zwicky, Arnold & Geoff Pullum (1983). Cliticization vs. inflection: English *n't*. *Language* 59: 502-513.

Theresa Biberauer
Department of Theoretical &
Applied Linguistics, MML
Sidgwick Avenue
Cambridge, CB3 9DA

mtb23@cam.ac.uk

Ian G. Roberts
Department of Theoretical &
Applied Linguistics, MML
Sidgwick Avenue
Cambridge, CB3 9DA

igr20@cam.ac.uk