

## **Scope variation in contrastive hierarchies of morphosyntactic features**

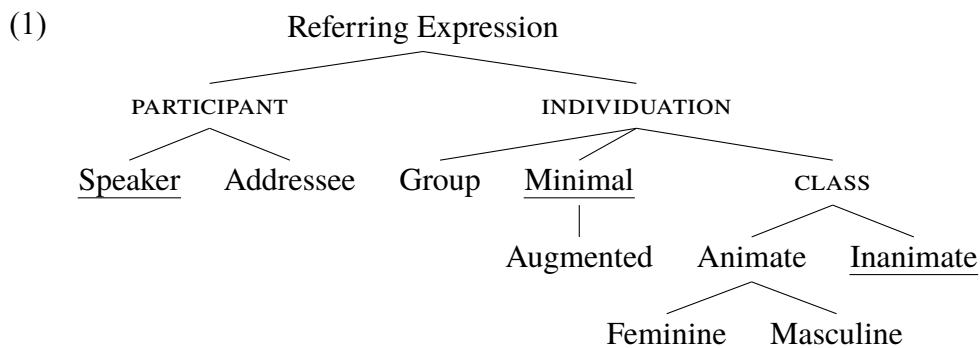
Elizabeth Cowper (University of Toronto) and Daniel Currie Hall (Saint Mary's University)

Morphosyntactic features exhibit hierarchical dependencies, in which the presence of one feature implies that of another. Following Bonet (1991) and Harley (1994), among others, these dependencies have often been represented using feature geometries like those used in autosegmental phonology (e.g., Clements and Hume 1995). However, the explanatory value of morphosyntactic feature geometries has been questioned (Harbour 2011, 2016; Harbour and Elsholtz 2012). We propose that another kind of hierarchical structure, also borrowed from phonology, serves as a useful model for morphosyntax: the contrastive hierarchy (Dresher 2009, 2015, 2016, this volume). Treating morphosyntactic hierarchies as contrastive hierarchies, we argue, can explain feature dependencies and their cross-linguistic variability as products of a general mechanism for acquiring contrasts from language-particular input (Cowper and Hall 2014). For the specific case of grammatical person, we further show that adopting the contrastive-hierarchy approach makes it possible to capture Harbour's (2016) typological generalizations while reducing the formal complexity of the features involved: the necessary features can be represented as familiar first-order predicates rather than as functions that add or subtract entities from a semilattice. Dresher's (2009) procedure for building contrastive hierarchies also implies a possible learning path; adopting this approach may thus make interesting predictions about the acquisition of morphosyntactic contrasts.

### **Feature geometries and their faults**

Morphosyntactic feature geometries can express dependency relations and provide a

useful visualization of what feature combinations are possible. For example, (1) shows the geometry proposed by Harley and Ritter (2002: 486) for  $\phi$ -features (person, number, and gender). Underlining indicates default values; e.g., Speaker is the default interpretation of a PARTICIPANT node with no marked dependent features.



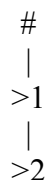
In this tree, the dependency of Feminine and Masculine on Animate encodes the fact that only animate nominals exhibit gender contrasts; likewise, Speaker and Addressee, which refer to specific discourse participants, each entail the presence of the more general PARTICIPANT. In addition to restricting the possible combinations of features through entailments of this sort, the geometry also provides a visual representation of specificity, which is relevant for vocabulary insertion in Distributed Morphology (Halle and Marantz 1993): a vocabulary item that spells out a subordinate feature in the tree is more specific than one that spells out any feature or node that dominates it, and will thus take precedence over it in the competition for insertion.

However, the motivation for feature geometries in morphosyntax is not as robust as in phonology. Feature geometries in phonology are supported by processes of spreading or delinking that operate on non-terminal nodes in the dependency structure. For example, homorganic nasal assimilation in various languages can be analyzed as spreading of the Place node to the nasal from an immediately following consonant

(Clements and Hume 1995: 270–271), taking with it all features dominated by Place (potentially Labial, Coronal, Dorsal, [ $\pm$ anterior], etc.) but leaving the manner and voicing of the nasal unaffected. No such support is found in syntax for morphosyntactic feature geometries: their subtrees do not move independently as syntactic constituents, nor can syntactic agreement reasonably be represented as spreading, sensitive to adjacency on a tier, as is phonological assimilation. Likewise, Béjar’s (2003: 77, 81) proposal for feature deletion in complex agreement systems simply deletes all marked features, rather than delinking proper subtrees of the geometry.

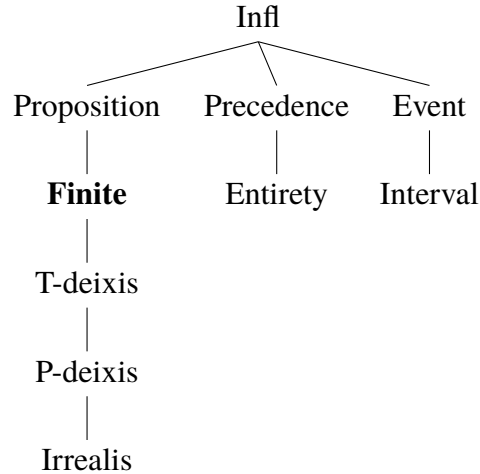
The motivation for geometric dependencies in morphosyntax has also been challenged, notably by Harbour (2011, 2016) and Harbour and Elsholtz (2012). Some are derivable from semantic entailment, and thus redundant. For example, in Cowper’s (2005b: 446) geometry for grammatical number features, shown in (2), the dependency of  $>2$  (which distinguishes plural from dual) on  $>1$  (which distinguishes dual and plural from singular) is mathematically inherent in the denotations of the features themselves:

(2) Number feature geometry from Cowper (2005b: 446)



Béjar (2003: 44) also explicitly takes feature-geometric trees to represent entailments. In such cases, the geometry merely represents a logical necessity; it may be a convenient tool for visualizing dependencies, but contributes nothing substantive of its own. Other feature-geometric dependencies, Harbour and Elsholtz (2012) argue, are pure stipulations, encoding unexplained observations. For example, consider the position of Finite in (3):

(3) Tense/mood/aspect feature geometry from Cowper (2005a: 14)



Finite is a purely syntactic feature with no semantic content, representing the presence of subject case and  $\phi$ -feature agreement. Its dependence on Proposition in (3) expresses, but does nothing to explain, the generalization that all finite clauses denote propositions (but not all propositional clauses are finite). Whereas the dependencies in (2) are redundant, derivable from the semantic content of the features themselves, the position of Finite in (3) is a stipulation. In neither instance does the feature geometry contribute directly to explaining the patterns it encodes.

Finally, the particular feature-geometric approach to person proposed by Harley and Ritter (2002) raises an empirical problem. In languages without a clusivity distinction, Speaker is the default interpretation of a bare PARTICIPANT node (Harley and Ritter 2002: 486); i.e., it is not a marked feature of first-person pronouns. However, languages with a clusivity distinction use both Speaker and Hearer as marked dependents of PARTICIPANT. Exclusive first-person forms are specified with Speaker, second-person forms with Hearer, and inclusive forms with both Speaker and Hearer. Harley and Ritter (2002: 490) propose that “[t]he learner can deduce that Speaker is not underspecified in her language from the presence of this inclusive/exclusive contrast.” The problem with

this approach is that such languages make no use of a bare PARTICIPANT node, contra Harley and Ritter's (2002: 509) expectation "that if a language has a pronoun with a complex geometry, the simpler geometries that form the subconstituents of the complex geometry are also available in that language." But if a bare PARTICIPANT node were available in languages with four-way person systems, we would (wrongly) expect to find pronouns referring to an undifferentiated participant both in these languages and in those that make only a participant/non-participant contrast, but not in languages with a three-way system (where a bare PARTICIPANT node is interpreted as specifically first person).

### **The contrastive-hierarchy approach**

Instead of feature geometries, then, we propose that morphosyntactic features are organized into contrastive hierarchies of the sort used in phonology by Dresher (2009) and others cited therein. Unlike a feature geometry, a contrastive hierarchy is not a subsegmental constituent structure; rather, it expresses the relative contrastive scope of features in the inventory as a whole.

#### *Contrastive hierarchies in phonology*

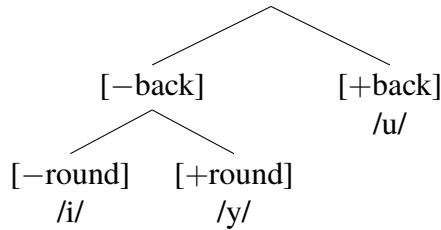
In phonology, contrastive hierarchies have been used as a mechanism for generating underspecified representations for segments: each phoneme is assigned enough features to distinguish it from the other phonemes with which it contrasts, but redundant features are omitted. Dresher (2009: 16) provides an explicit procedure for constructing a contrastive hierarchy by using features to divide a phonological inventory until each phoneme has a unique representation; this is the Successive Division Algorithm (SDA), shown in (4).

- (4) Successive Division Algorithm (SDA; Dresher 2009: 16)

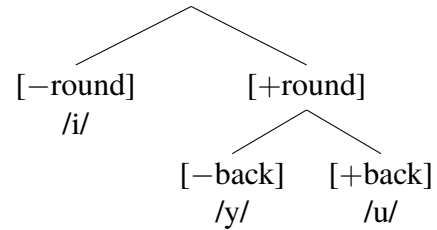
- a. Begin with *no* feature specifications: assume all sounds are allophones of a single undifferentiated phoneme.
- b. If the set is found to consist of more than one contrasting member, select a feature and divide the set into as many subsets as the feature allows for.
- c. Repeat step (b) in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.

The SDA is neutral as to whether features are selected from a universal set (Jakobson, Fant, and Halle 1952 and much subsequent work), or induced by learners from the primary linguistic data (e.g., Mielke 2008). Even if the set of features is universal, though, their hierarchical ordering can vary from one language to another, so that languages with phonetically similar inventories may use different sets of feature specifications. For example, consider languages with three high vowels /i y u/, like Finnish or French. If the features [±back] and [±round] are used to differentiate these phonemes, the SDA allows two different sets of feature specifications. Whichever feature is used first will be assigned to all three vowels, dividing them into one subset of two and one singleton. The second feature will divide the subset of two, but will not be assigned to the third vowel, because that vowel has already been distinguished from the others by the first feature. The two resulting hierarchies are shown in (5) (adapted from Burstynsky 1968: 11):

(5) a.  $[\pm\text{back}] \gg [\pm\text{round}]$



b.  $[\pm\text{round}] \gg [\pm\text{back}]$



As Burstynsky argues, the phonological behaviour of these vowels in Quebec French indicates that the language uses the hierarchy in (5a): the high front vowels /i/ and /y/ pattern together as a natural class in triggering assibilation of dental stops. In Finnish, however, (5b) appears to be the correct ordering: while the round vowels /y/ and /u/ participate in place harmony, /i/ is transparent to it, behaving as though it is unspecified for  $[\pm\text{back}]$  (Hall 2017; see also Mackenzie 2011: §3 for another example of cross-linguistic variation in feature scope).

The SDA guarantees that no more features will be used than are required to differentiate the contrasting elements in the inventory. It thus limits the features that can be assigned to any given inventory, while still allowing for cross-linguistic variation within those limits. Contrastive hierarchies also offer insight into phonetic enhancement and the typology of phonemic inventories (Hall 2011), patterns of reduction (Spahr 2014), and diachronic change (Dresher et al. 2014).

#### *Contrastive hierarchies in morphosyntax*

We propose that the dependency relations among interpretable morphosyntactic features reflect contrastive hierarchies, not feature-geometric structures. The central insight of a contrastive hierarchy is that the applicability and the interpretation of a feature depend on the domain in which it is contrastive, as defined by the features above it in the hierarchy. Both feature geometries and contrastive hierarchies provide a way to describe scope

relations among features. However, the expressive possibilities made available by feature geometry differ from those that can be expressed by a contrastive hierarchy. In a feature geometry with privative features, only marked features can have dependents. Suppose that a feature F has two dependents, G and H. If H is a dependent of G, giving a non-branching geometry like the one in (2), then there are three possible representations, listed in (6a). If H and G are sisters in the dependency structure, both dependent on F but neither of the two on the other, then there are four possible representations, listed in (6b). There is no way, without specifying  $[-G]$  explicitly, to restrict the domain of  $[H]$  to only those instances of F that do not bear  $[G]$ ; i.e., to rule out the combination FGH in (6b).

- (6) a.  $\{F, FG, FGH\}$                       b.  $\{F, FG, FH, FGH\}$

With a contrastive hierarchy, the assignment of features divides the inventory, rather than structuring the representation, and a given feature takes scope only over the sub-inventory that it divides. This means that there are three possible contrastive hierarchies in which a set defined by the feature  $[F]$  is divided first by  $[G]$  and then by  $[H]$ . In all three cases,  $[G]$  divides the inventory of  $[F]$ -bearing elements into those bearing  $[G]$  and those lacking  $[G]$ . This gives two sub-inventories, specified  $[F]$  and  $[FG]$ .  $[H]$  can then divide either one or both of these sub-inventories. If  $[H]$  divides only the  $[FG]$  set, the result is the same as in (6a). If  $[H]$  divides both sub-inventories, the result is the same as (6b). The third possibility, though, is that  $[H]$  divides the  $[F]$  set but not the  $[FG]$  set; this yields three sub-inventories  $\{F, FG, FH\}$  in which  $[G]$  and  $[H]$  do not co-occur.

If features are privative, as assumed in most feature-geometric approaches, then contrastive hierarchies allow more different dependency structures than feature



geometries do. The difference in expressive power between the two approaches is less obvious with binary features, but we show below that in the specific case of person features, contrastive hierarchies allow an account of the typological patterns using simpler features than those proposed by Harbour (2016), while retaining the key advantages of his approach. The question of whether features in general are binary, privative, or a mix of the two is a separate issue (on which see Cowper and Hall 2014); in all cases, contrastive hierarchies offer a principled, non-stipulative way of representing dependencies, and their expressive power can reduce the burden on the definitions of the features themselves.

### **The representation and typology of person**

As just stated, the notion of contrastive scope is not tied to any particular conception of what a feature is. Contrastive hierarchies can be combined with features with various formal properties. However, in at least one case, the contrastive-hierarchy approach makes it possible to use a less powerful kind of feature.

#### *Harbour's person features*

Harbour (2016) presents a comprehensive theory of grammatical person that accounts for the attested typological range of systems of contrast, but requires features to be formalized as operators that add or subtract elements of semilattices, rather than as first-order predicates. The features operate on a universal person ontology comprising a unique speaker *i*, a unique hearer *u*, and arbitrarily many others *o*. From an extensive typological survey, Harbour observes that only five of the 15 logically possible sets of person contrasts ('partitions') are attested. These are shown in (7); in Harbour's notation, a subscript *o* indicates the addition of zero or more others (third persons):

(7) a	‘Monopartition’	no contrasts	$\{i_o, iu_o, u_o, o_o\}$
b	Author bipartition	first vs. non-first	$\{i_o, iu_o\} / \{u_o, o_o\}$
c	Participant bipartition	non-third vs. third	$\{i_o, iu_o, u_o\} / o_o$
d	Standard tripartition	1st vs. 2nd vs. third	$\{i_o, iu_o\} / u_o / o_o$
e	Quadripartition	excl. vs. incl. vs. 2nd vs. 3rd	$i_o / iu_o / u_o / o_o$

Other logically possible sets of contrasts are unattested; for example, no language is known to make only an addressee bipartition (second vs. non-second). Such a partition may arise as a syncretism in a particular morphological paradigm, but only in a language with a richer overall system of person contrasts (tripartition or quadripartition).

The fact that the most richly articulated system has quadripartition suggests that there are no more than two binary features available for marking person contrasts. The existence of author and participant bipartitions suggests that these features are [ $\pm$ author] and [ $\pm$ participant], and the absence of addressee bipartitions suggests that [ $\pm$ hearer] is not available, at least not as the only feature in a system. The challenge, then, is to explain how tripartition and quadripartition, each of which requires two features, can exist as distinct types of systems without positing features that would also generate unattested partitions.

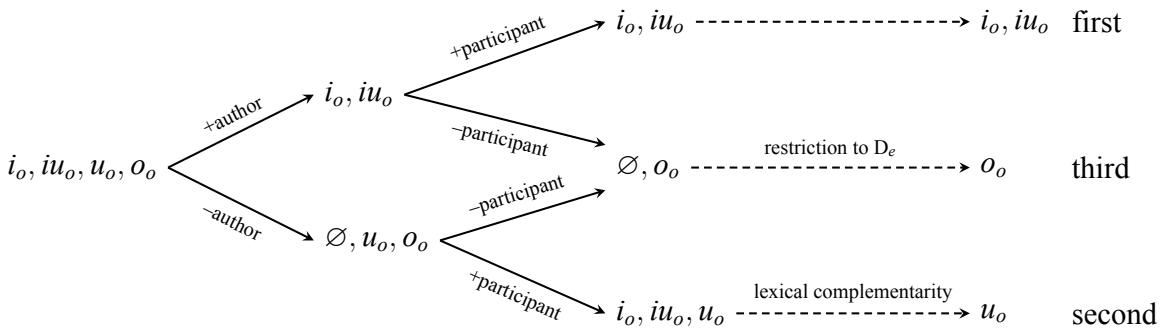
Like others, notably Halle (1997), Harbour posits that UG provides two binary person features, [ $\pm$ author] and [ $\pm$ participant]. Harbour’s crucial innovation is that his features, rather than denoting first-order predicates such as ‘includes the speaker’ or ‘includes a discourse participant,’ are functions that operate on (semi)lattices to add or subtract individuals. The effects of his features are as follows:

- [ $\pm$ author] adds the speaker  $i$  to a lattice.

- [-author] subtracts the speaker  $i$  from a lattice.
- [+participant] disjointly adds all discourse participants  $\{i, iu, u\}$  to a lattice.
- [-participant] subtracts all participants  $\{i, iu, u\}$  from a lattice.

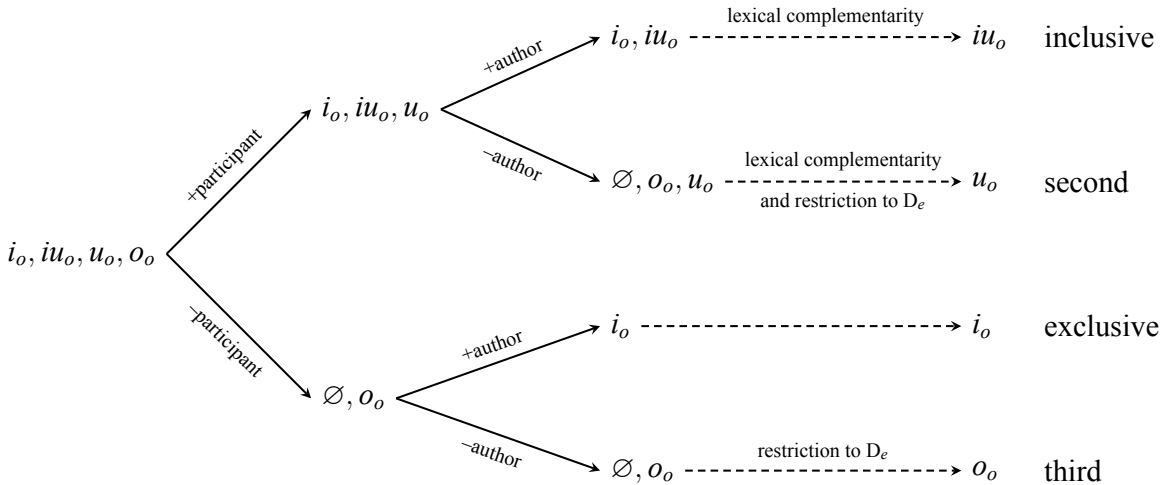
As in other accounts, monopartition systems use no person features, and each of the two attested bipartitions uses one feature. The importance of Harbour's formal implementation of the features as operations emerges in his account of tripartition and quadripartition. Each of these two systems uses both features, but in opposite orders. Applying  $[\pm\text{author}]$  before  $[\pm\text{participant}]$  derives the standard tripartition as in (8).

(8)  $[\pm\text{author}]$  before  $[\pm\text{participant}]$  (Harbour 2016: 99)



Applying  $[\pm\text{participant}]$  before  $[\pm\text{author}]$  derives quadripartition as in (9).

(9)  $[\pm\text{participant}]$  before  $[\pm\text{author}]$  (Harbour 2016: 99)



In addition to the redefinition of the features themselves as operations that can

apply in a particular order, Harbour requires two interpretive principles that apply to their output. The first, restriction to the domain of entities ( $D_e$ ), excludes the empty set  $\emptyset$  from the output. In both (8) and (9), the sequences of features that generate third person yield semilattices that include the empty set as well as all non-empty sets of non-participants; restriction to  $D_e$  ensures that only non-empty third persons are included.

The second principle, lexical complementarity, eliminates overlap between subsets in accordance with the Elsewhere Principle. In (8), for example, because the output of  $[-\text{author}, +\text{participant}]$ ,  $\{i_o, iu_o, u_o\}$ , is a superset of the output of  $[+\text{author}, +\text{participant}]$ ,  $\{i_o, iu_o\}$ , lexical complementarity restricts  $[-\text{author}, +\text{participant}]$  to  $u_o$  (its only member that is not also in  $[+\text{author}, +\text{participant}]$ ). Likewise, the sequence  $[+\text{participant}, +\text{author}]$  in (9) yields  $\{i_o, iu_o\}$ , which would be a general first person rather than a first person inclusive; its interpretation is restricted to the inclusive ( $iu_o$ ) because  $[-\text{participant}, +\text{author}]$  yields  $i_o$ . Lexical complementarity is similar to, but must apply independently from, other instances of the subset principle. For example, the fact that a form like English *you* in a tripartition system is interpreted as not including the speaker cannot be merely a scalar implicature, nor can it be derived from competition between vocabulary items in Distributed Morphology.

Harbour's approach generates exactly the attested range of systems of grammatical person contrasts, but it is formally complex. The features themselves must be defined as operations rather than as predicates, and their output must be subject to further rules. Additionally, in the derivation of tripartition in (8), third person has two possible representations:  $[-\text{author}, -\text{participant}]$  or  $[+\text{author}, -\text{participant}]$ . Harbour (2016: 92–93) posits a parameter to allow tripartition languages to use one or the other,

but not both.

The diagrams in (8) and (9) are not contrastive hierarchies like those generated by the SDA using traditional first-order predicate features. Notably, they are not trees, as (8) contains two routes from  $\{i_o, i_{u_o}, u_o, o_o\}$  to third person. However, the sequencing of operations in Harbour's approach is broadly analogous to the role of contrastive scope in the SDA: the first feature defines the formal objects to which the second applies, which in turn determine the interpretive consequences of the second feature. We propose that the key insights of Harbour's approach can be retained with formally simpler features if the ordering of features is recast as scope-taking in a contrastive hierarchy.

#### *Simplifying the person features*

In general, interpretable morphosyntactic features, like their phonological counterparts, have been understood as first-order predicates, denoting a property [F] that a given lexical item either has (represented by [+F] in a binary system, or by [F] if the feature is privative) or lacks (represented as [-F] or the absence of privative [F]). When a lexical item is specified with more than one such feature, the two compose intersectively. In the absence of the scope differences made possible by the contrastive hierarchy, this means that the order of application of features should make no difference to the result. However, the adoption of the contrastive-hierarchy approach can create a situation in which order makes a difference.

We assume Harbour's (2016) ontology of persons, repeated in (10), where  $i$  is the unique author,  $u$  the unique addressee, and  $o, o', o''$ , etc. an arbitrary number of others.

$$(10) \quad \pi = \{i, u, o, o', o'', \dots\}$$

The inventory to be divided is the set of possible combinations of persons; in

other words, the power set of  $\pi$ .<sup>1</sup>

Like Harbour, we posit two binary features,  $[\pm\text{author}]$  and  $[\pm\text{participant}]$ .

However, we propose that they denote first-order predicates as in (11).

- (11) a.  $[\text{+author}]$  = ‘includes the speaker’  
b.  $[\text{-author}]$  = ‘does not include the speaker’  
c.  $[\text{+participant}]$  = ‘includes a(t least one) discourse participant’  
d.  $[\text{-participant}]$  = ‘does not include a discourse participant’

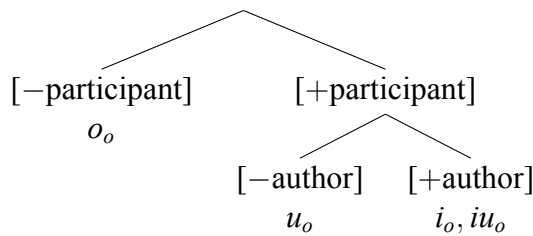
The person system of a given language may use one, both, or neither of the two features. If a language uses both, then one of the features will take scope over the other. Crosslinguistically, either order is possible. This gives five possible situations, the first three of which are listed in (12).

- (12) a. The language uses no person features. No person-based distinctions are found in the grammar. This is monopartition, as in (7a).  
b. The language uses only  $[\pm\text{author}]$ . First persons (both inclusive and exclusive) are distinguished from all others. There is no clusivity contrast, and second persons do not contrast with third persons. This yields the author bipartition (7b), dividing  $[\text{-author}]$  persons ( $u_o, o_o$ ) from  $[\text{+author}]$  persons ( $i_o, iu_o$ ).  
c. The language uses only  $[\pm\text{participant}]$ . First and second persons are together distinguished from third persons. There is no clusivity contrast, and no contrast between first and second persons. This is the participant bipartition in (7c), which separates  $[\text{+participant}]$  persons ( $i_o, u_o, iu_o$ ) from  $[\text{-participant}]$  persons ( $o_o$ ).

If the language uses both features, then their relative scope determines how the

inventory is partitioned. If [ $\pm$ participant] takes wider scope, then an initial division is made between [+participant] first and second persons, on the one hand, and [-participant] third persons on the other. Then, [ $\pm$ author] makes a second division among the [+participant] members, separating those that include *i* from those that do not. This gives the standard tripartition (7d), with contrasting first, second, and third persons but no clusivity distinction, as shown in (13).

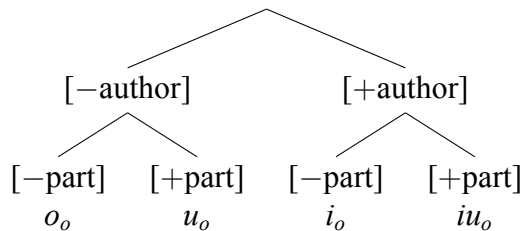
(13) Tripartition: [ $\pm$ participant] takes scope over [ $\pm$ author]



The final possibility has [ $\pm$ author] taking wider scope, making the first division between all first persons (inclusive or exclusive) on the one hand, and second and third persons on the other. Then [ $\pm$ participant] divides each of the subinventories. The division of the [-author] subinventory is straightforward: all elements including *u* are [+participant], and all those that lack *u* are [-participant]. The division of the [+author] subinventory is slightly less obvious, since at first blush, all of its members are [+participant] as defined in (11c). We propose that in this instance, the interpretation of [ $\pm$ participant] is automatically narrowed to ‘{includes, does not include} a participant other than speaker’. Essentially, the effect is to reinterpret [ $\pm$ participant] as referring only to the addressee. This is the only possible interpretation that allows it to be contrastive over an inventory where the inclusion of the speaker has already been marked. Note that this same narrowing, while not strictly necessary, does no harm if it applies to the division of the [-author] subinventory by [ $\pm$ participant].

Dividing the [+author] subinventory with the more narrowly defined [±participant] gives the clusivity distinction, deriving (7e) as in (14).

(14) Quadripartition: [±author] takes scope over [±participant]



One might argue that this approach smuggles the feature [±hearer] into the set of possible person systems, making the account vulnerable to Harbour’s (2016: §8.2) objections. He persuasively argues against approaches that involve parametric choice either between the use of [±participant] and [±hearer] or between allowing and excluding the combination of values that would distinguish inclusive from exclusive first persons (e.g., Halle 1997). While such systems maximally distinguish the same four possible persons as do Harbour’s features and the ones proposed here, the logical combinations of the three binary features predict more than the five attested partitions. Specifically, they do not explain why [±hearer] is never used alone, distinguishing second persons from an undifferentiated category including first and third persons. In addition, they do not explain why, if one combination of features can be ‘parametrically deactivated,’ others cannot.

We avoid these objections by linking the parametric variation in the semantics of [±participant] to its contrastive scope. A given feature’s interpretation is consistently contingent on the domain in which it marks a contrast. This is analogous to the phonetic interpretation of phonological features. For example, consider a vowel inventory specified first with [±high], and then within each resulting subinventory for [±back]. The



articulatory and acoustic difference between [–back] /i/ and [+back] /u/ in the [+high] subinventory is likely to be appreciably greater than that between [–back] /e/ and [+back] /o/ in the [–high] subinventory; the shape of the oral cavity dictates that the phonetic distance between [–back] and [+back] narrows as the height of the tongue decreases. Similarly, Clements (1991) represents degrees of vowel height by successive hierarchical application of a single feature [±open].<sup>2</sup> A vowel marked [+open] at the highest division in the hierarchy is low relative to the phonetic space as a whole. Specifying [+open] within the [–open] branch picks out the lowest vowels within the non-low category, namely mid vowels. While [±open] consistently indicates a single dimension of phonetic contrast, the difference along that dimension is wider or narrower according to the feature’s place in the hierarchy. Although the meaning of [±participant] is not gradient as vowel features are, its interpretation similarly depends on its position in the contrastive hierarchy.

The proposed features thus give exactly the attested set of person systems, with no need to invoke lexical complementarity: each combination of features straightforwardly denotes the appropriate set of possible referents, not a proper superset. The interpretive narrowing of [±participant] in (14), which gives essentially [±hearer], derives from its position in the contrastive hierarchy. No analogous narrowing of [±author] is possible to allow full cross-classification in (13), because the interpretation ‘speaker other than a discourse participant’ is nonsensical; this ordering of the features produces only the standard tripartition. We conclude that with a contrastive hierarchy, it is possible to account for exactly the attested person partitions with first-order features and no more additional machinery than Harbour (2016) requires.

## Acquisition

In addition to ensuring that representations include only contrastive features, the SDA suggests a learning path: as children acquire contrasts, they build the representations that encode them (Dresher 2014: §4). Ideally, we might expect the order of acquisition to correspond to hierarchical scope: features that are higher in the tree would be acquired first.

However, actual acquisition paths will likely turn out to be more complicated. Some proposed phonological contrastive hierarchies do not map readily to acquisition sequences, and suggest that learners may need to do some backtracking. For example, Hall (2007) proposes a hierarchy for Czech consonants in which the first two features divide the inventory into sonorants, obstruents, and the trilled fricative /ř/. /ř/ is typically the *last* consonant Czech children accurately produce, though this could plausibly be attributed to its articulatory complexity rather than to its phonological encoding.

In the most straightforward mapping from contrastive scope to order of acquisition, our person feature hierarchies make the following predictions. In a tripartition language, where [ $\pm$ participant] takes wider scope, children will first distinguish participants from third persons. Early learners may fail to distinguish the representations of first and second person, and thus seem to confuse first- and second-person forms. In a quadripartition language, with [ $\pm$ author] above [ $\pm$ participant], children will begin by distinguishing first persons from second and third; early learners may conflate second person with third, and inclusive with exclusive.

To what extent are these predictions borne out? It is difficult to say, partly because of an abundance of potential confounds, and partly because investigations into

the acquisition of person systems seldom probe the question of contrasts directly.

Children acquiring tripartition languages do sometimes confuse first and second person in production, but second person is often the first to be mastered in comprehension; see, e.g., Moyer et al. (2015: 2) and references cited therein, especially Oshima-Takane (1992). Considerably less work has been done on the acquisition of quadripartition languages.

Although the prospect is complicated by mismatches between comprehension and production and by the possibility of back-tracking, the predictions for acquisition made by the contrastive-hierarchy approach show promise. Moreover, this view of how features are organized has broader implications for how questions about acquisition should be framed: to look at how learners acquire person or any analogous grammatical system, we should focus on acquisition of *distinctions* (à la Jakobson 1941) rather than of *items*.

## **Conclusion**

In phonology, feature geometries are motivated not only by dependency relations among features, but by the fact that those dependencies are active in autosegmental processes (spreading and delinking). In other words, there is evidence that they are part of the structural representation of each segment. It remains open exactly how phonological contrastive hierarchies relate to phonological feature geometries, since they encode some of the same information in different ways. (See Iosad (2012) for a proposal combining the SDA with Morén's (2003, 2006) Parallel Structures Model of feature geometry.)

Broadly speaking, contrastive hierarchies are paradigmatic, defining systems of oppositions, and feature geometries are syntagmatic, structuring combinations of features

in phonological representations.

In morphosyntax, syntagmatic representations are phrase structure (and word structure) trees, and operations like Merge and Move apply to them. No clear evidence has yet emerged to suggest that the dependencies among morphosyntactic features are relevant to any of these operations. The dependencies that have previously been expressed in morphosyntactic feature geometries are real, but they are paradigmatic rather than syntagmatic.

Contrastive hierarchies offer a way of representing these dependencies that does not imply that morphosyntactic features should engage in the same kinds of spreading and delinking operations that apply to phonological autosegments. At the same time, contrastive hierarchies do other useful things that feature geometries cannot. They represent scope relations between features in a way that is neither redundant nor stipulative, and account for the fact that the interpretation of a feature depends in part on the domain in which it is contrastive. They are compatible with multiple views of the formal properties of features themselves, and in the case of privative features make it possible to express the dependence of one feature on the absence of another. For person features, we have shown how a contrastive-hierarchy approach can reduce the required formal complexity of the features themselves. Finally, they present an opportunity to shed new light on the acquisition of grammatical elements by framing the question as pertaining to contrasts rather than to vocabulary items.

## **References**

Béjar, Susana. 2003 *Phi-syntax: A theory of agreement*. Ph.D. thesis, University of Toronto.

- Bonet, Eulàlia. 1991. *Morphology after syntax: Pronominal clitics in Romance*. Ph.D. thesis, Massachusetts Institute of Technology.
- Burstynsky, Edward N. 1968. Quelques observations de phonologie générative appliquées au français canadien. In *Recherches sur la structure phonique du français canadien*, ed. Pierre R. Léon, 9–17. Montréal: Marcel Didier.
- Clements, G. N. 1991. Vowel height assimilation in Bantu languages. *BLS* 17: 25–64.
- Clements, G. N., and Elizabeth Hume. 1995. The internal organization of speech sounds. In *The handbook of phonological theory*, ed. John Goldsmith, 245–306. Oxford: Blackwell.
- Cowper, Elizabeth. 2005a. The geometry of interpretable features: Infl in English and Spanish. *Language* 81(1): 10–46.
- Cowper, Elizabeth. 2005b. A note on number. *Linguistic Inquiry* 36(3): 441–455.
- Cowper, Elizabeth, and Daniel Currie Hall. 2014. Reductiō ad discrīmen: Where features come from. *Nordlyd* 41(2): 145–164.
- Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
- Dresher, B. Elan. 2014. The arch not the stones: Universal feature theory without universal features. *Nordlyd* 41(2): 165–181.
- Dresher, B. Elan. 2015. The motivation for contrastive feature hierarchies in phonology. *Linguistic Variation* 15(1): 1–40. DOI: 10.1075/lv.15.1.01dre.
- Dresher, B. Elan. 2016. Contrast in phonology, 1867–1967: History and development. *Annual Review of Linguistics* 2: 53–73.

- Dresher, B. Elan. (this volume) Contrastive feature hierarchies in phonology: Universality and variation.
- Dresher, B. Elan, Christopher Harvey, and Will Oxford. 2014. Contrast shift as a type of diachronic change. In *NELS 43: Proceedings of the forty-third annual meeting of the North East Linguistic Society*, vol 1, ed. Hsin-Lun Huang, Ethan Poole and Amanda Rysling, 103–116. Amherst, MA: GLSA.
- Hall, Daniel Currie. 2007. *The role and representation of contrast in phonological theory*. Ph.D. thesis, University of Toronto.
- Hall, Daniel Currie. 2011. Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology* 28(1): 1–54.
- Hall, Daniel Currie. 2017. Contrastive specification in phonology. In *Oxford Research Encyclopedia of Linguistics*, ed. Mark Aronoff. Published online. DOI: 10.1093/acrefore/9780199384655.013.26.
- Halle, Morris. 1997. Distributed morphology: Impoverishment and fission. *MIT Working Papers in Linguistics* 30: 125–149.
- Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, MA: MIT Press.
- Harbour, Daniel. 2011. Descriptive and explanatory markedness. *Morphology* 21(2): 223–245.
- Harbour, Daniel. 2016. *Impossible persons*. Cambridge, MA: MIT Press.
- Harbour, Daniel, and Christian Elsholtz. 2012. Feature geometry: Self-destructed.

- Ms., Queen Mary University of London and Technische Universität Graz.
- Harley, Heidi. 1994. Hug a tree: Deriving the morphosyntactic feature hierarchy. *MIT Working Papers in Linguistics* 21: 289–320.
- Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3): 482–526.
- Iosad, Pavel. 2012. *Representation and variation in substance-free phonology: A case study in Celtic*. Ph.D. thesis, University of Tromsø.
- Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Uppsala Universitet.
- Jakobson, Roman, Gunnar Fant, and Morris Halle. 1952. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
- Mackenzie, Sara. 2011. Contrast and the evaluation of similarity: Evidence from consonant harmony. *Lingua* 121(8): 1401–1423.
- Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.
- Morén, Bruce. 2003. The Parallel Structures Model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15: 194–270.
- Morén, Bruce. 2006. The division of labor between segment-internal structure and violable constraints. In *Freedom of analysis?*, ed. Sylvia Blaho, Patrik Bye, and Martin Krämer, 313–344. The Hague: Mouton.
- Moyer, Morgan, Kaitlyn Harrigan, Valentine Hacquard, and Jeffrey Lidz. 2015. 2-year-olds' comprehension of personal pronouns. In *BUCLD 39 online*

*proceedings supplement*, ed. Elizabeth Grillo, Kyle Jepson, and Maria

LaMendola, 11 pp. Boston: Boston University. Published online at

<http://www.bu.edu/buclid/supplementvol39/>.

Oshima-Takane, Yuriko. 1992. Analysis of pronominal errors: A case-study.

*Journal of Child Language* 19(1): 111–131.

Spahr, Christopher. 2014. A contrastive hierarchical account of positional

neutralization. *The Linguistic Review* 31(3–4): 551–585. DOI: 10.1515/tlr-

2014-0008.

---

We are grateful to Neil Banerjee, Bronwyn Bjorkman, B. Elan Dresher, David Lightfoot, and Andrew Peters, and the audiences at the 2016 Tromsø Hierarchies Workshop, the 2017 Georgetown University Round Table, and the 2017 Annual Meeting of the Canadian Linguistic Association, for helpful questions and discussion on this topic.

<sup>1</sup> Or perhaps only the non-empty members of  $P(\pi)$ , as per Harbour's operation of restriction to  $D_e$ . Note, however, that for us this would be a restriction on the input to the SDA rather than a repair on its output, and that in any case both our system and Harbour's predict (plausibly enough) that if empty persons are conceptually possible at all (cf. Harbour 2016: 85–86), they will be referred to with the same forms as third persons ( $o_o$ ).

<sup>2</sup> This re-application of (potentially contrary values of) a single feature is similar to Harbour's (2011, 2014) treatment of grammatical number. For example, applying first [–minimal] and then [+minimal] picks out the minimal sub-region of the non-minimal region of the ontological space.