Person and number in pronouns: a feature-geometric analysis

Heidi Harley, University of Arizona
Elizabeth Ritter, University of Calgary

Contact info:

Heidi Harley
Douglass 200E
Department of Linguistics
University of Arizona
Tucson, AZ 85721-0028

(520)626-3554
fax: (520)626-9014
hharley@u.arizona.edu

Elizabeth Ritter
Department of Linguistics
University of Calgary
2500 University Drive NW
Calgary, AB T2N 1N4
CANADA

(403) 220-7315
ritter@ucalgary.ca

Abstract

The set of person and number features necessary to characterize the pronominal paradigms of the world's languages is highly constrained, and their interaction is demonstrably systematic. Here, we develop a geometric representation of morphosyntactic features which provides a principled explanation for the observed restrictions on these paradigms. The organization of this geometry represents the grammaticalization of fundamental cognitive categories, such as reference, plurality, and taxonomy. We motivate the geometry through the analysis of pronoun paradigms in a broad range of genetically distinct languages.[*]

Introduction. It is generally accepted that syntactic and phonological representations are formal in nature and highly structured. Morphology, however, is often seen as a gray area in which amorphous bundles of features connect phonology with syntax via a series of ad hoc correspondence rules. Yet it is clear from the pronoun and agreement paradigms of the world's languages that Universal Grammar provides a highly constrained set of morphological features, and moreover that these features are systematically and hierarchically organized.[1] In this paper we develop a structured representation of person and number features intended to predict the range and types of interactions among them. More specifically, we shall motivate the claims in 1:

(1) Claims

a. The language faculty represents pronominal elements with a geometry of morphological features.

b. The organization of this geometry is constrained and motivated by conceptual considerations.

c. Cross-linguistic variation and paradigm-internal gaps and syncretisms are constrained by the hierarchical organization of features in the universal geometry.

d. The interpretation of sub-trees of the geometry may be relativized in tightly constrained ways so that language-specific interpretation of a given feature will depend in part upon the contrasts available within that language.

In order to substantiate these claims we demonstrate that a geometric analysis can shed light on children's acquisition of pronouns, and on paradigms which manifest a range of unusual properties. In the course of this demonstration we draw on our own database of one hundred and ten languages, as well as descriptions of other exceptional languages discussed in the literature.

1. Feature Geometries

There are natural classes of morphological features. This is reflected in the universal classificatory use of the terms “person”, “number” and “gender,” as well as other classes of features not discussed here. At least this much organization is unconditionally assumed, and further organized relationships among these classes of features have often been noted, although not often treated theoretically. For instance, Greenberg (1963) observed a number of cross-linguistic generalizations concerning the clustering of features, describing, e.g., the dependence of gender on number (Universal 32: ‘Whenever a verb agrees with a nominal subject or object in gender it also agrees in number.’), and the dependence of dual number on plural number (Universal 34: ‘No language has a dual [number] unless it has a plural.’) (Greenberg 1963:94).

Although these dependencies must be the consequence of some aspect of Universal Grammar, morphological theory has not in general attempted to provide an account of them. It seems to be widely assumed that features are collected in unstructured bundles, and that morphological rules may freely refer to any feature or group of features, regardless of whether or not they form a natural class.

1.1 The problem with unstructured bundles of features. Many modern theories of morphology claim that morphosyntactic features are grouped in unstructured bundles, and even make a virtue of it — despite the fact that a certain amount of organization is implicit in the very terminology common to all morphological theories. For examples, consider the discussion of ‘Morphosyntactic Representations’ (MSRs) in Anderson (1992:92): ‘The minimal (and thus the most desirable) theory of MSRs … is one that would assign them no internal structure at all.’ Yet even given this explicit assertion, Anderson does not assume that MSRs representing agreement features are entirely unstructured. He implicitly assigns attribute-value structure to them: there is

at least a PERSON attribute which may bear the values {[±me], [±you]}; a NUMBER attribute which may bear the values {[±pl]}, and a GENDER attribute which may bear a series of familiar values such as {[±masc], [±fem], [±neut]}. This organization, whether implicitly assumed, as in Anderson (1992), or explicitly assumed, as in the LFG or HPSG frameworks, is essential to prevent the unmotivated combinatorial explosion of potential cross-classifications. In a system where a single form could bear any combination of the features [1], [2], [sg], [pl], [masc], [fem], there would be 2^6 possible combinations of feature values. Collecting the person, number and gender features together into mutually exclusive subgroups brings the paradigmatic possibilities down to something much more closely approximating the facts of natural language.[2, 3]

Even given these subgroups, usually considered to be so basic as not to merit discussion, the organization across subgroups that is implied by the Greenbergian generalizations listed above does not fall out of any account of morphosyntactic features that assumes that feature groups are collected in unstructured bundles. Noyer (1992, 1997) represents an important attempt to provide a theoretical framework which predicts the existence of universals such as those catalogued by Greenberg. He posits a Universal Feature Hierarchy (UFH), which interacts with the unordered bundles of features, to explain regularities captured in Greenberg's generalizations. A partial version of his UFH can be seen in 2:

(2) Noyer's (1992) Universal Feature Hierarchy

person features > number features > gender features > class features

However, under Noyer's treatment the feature groups themselves are still represented as unordered bundles. The UFH is intended as a constraint which determines, for a given bundle, which features may be deleted or otherwise manipulated. For example, a bundle with an

incompatible number and gender combination will be amended by deleting the offending gender feature as in 3, since gender is lower on the UFH than number.

(3) Filter: *[2 f]

[2 pl f] → [2 pl]

The formal mechanism of the deletion rules themselves, however, without an additional constraint like the UFH, makes no predictions about possible syncretisms or groupings, and is subject to the criticisms noted earlier.

Our goal is to develop representations from which these properties will naturally follow, without recourse to a separate stipulated metric. Borrowing heavily from the phonological model, we propose a morphological feature geometry which is intended to constrain pronoun and agreement systems, much as phonological feature geometries constrain sound systems. Such geometries have been discussed in detail in Harley (1994), Ritter (1997), and a similar geometry which also includes syntactic feature information was developed in Gazdar and Pullum (1982) and discussed in Gazdar et al. (1985).

1.2 Structuring the representation. Our feature geometry is of the same formal type as the phonological feature geometries originally proposed in Clements (1985) and Sagey (1986). The morphological feature geometry is motivated by the need for the same theoretical mechanisms as its phonological counterpart. The primary goals of a phonological feature geometry, as described by Noyer (1992), are outlined in 4:

(4) a. Subtrees define natural classes of features for phonological rules.

b. (Some) nodes in a feature tree correspond to the anatomy of the vocal tract.

c. The dependency relation encodes contrastiveness.

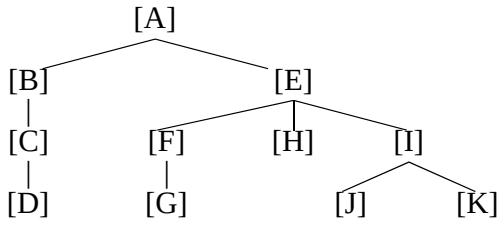
(Noyer 1992:46-47)

Harley (1994) showed that a morphological feature geometry could accomplish 4a and 4c, and also argued that the feature-geometric approach to morphology defines a notion of complexity from which the effects of Noyer's UFH can be derived.

Here, we also wish to maintain that there is a natural morphological analogue to 4b, that is, there are external reasons for the feature groupings which arise. It is undeniable that the organization of the vocal tract both constrains and motivates phonological feature organization. At first glance, it might seem that the groupings of morphological features, independent of physical restrictions, cannot be attributed to external factors. Following Ritter (1997), we claim that the external factors which play a role in morphology are conceptual in nature. Specifically, notions such as deixis, countability and taxonomy constrain and motivate the relationships which are apparent among morphological features and represented in the formal geometry. Subtrees of the geometry represent the grammaticization of natural cognitive categories, accounting for the apparent yet tenuous relationship between grammatical features and meaningful concepts.

1.3 The formal properties of feature geometries. Before discussing the details of our proposal, we briefly review some formal properties of feature geometries. Consider the abstract feature-geometric tree represented in 5.

(5)



Following Harley (1994) we assume that features are monovalent, and only appear if they have a positive value. That is, feature [B] will only appear in the structure if feature [B] is active. A feature [-B] does not occur.

In a feature geometry, it is generally the case that if one morphological feature is logically implied by another, this relation is captured through dependence. For instance, in (5), feature [G] entails that feature [F] is present because feature [G] is represented as dependent on feature [F]. Feature [F] may not be eliminated from the geometry without also eliminating feature [G].

It is worth noting, however, that the logical implication produced by the everyday interpretation of a given pair of features need not be represented in the geometry. For example, the 1st person inclusive is logically non-singular, as it necessarily includes the speaker and at least one addressee. However, not all languages mark inclusive forms as plural.[4] This shows that this logical implication is not morphologically relevant in all languages. (A phonological analogue might be the interdependence of [+sonorant] and [+voice]: although all sonorants are voiced, they are not represented as dependents of the [+voice] node). As we illustrate below, our geometry does not encode the implication between inclusive person and plural number as a dependence relation, and this correctly predicts the existence of rare languages that have an inclusive singular.[5]

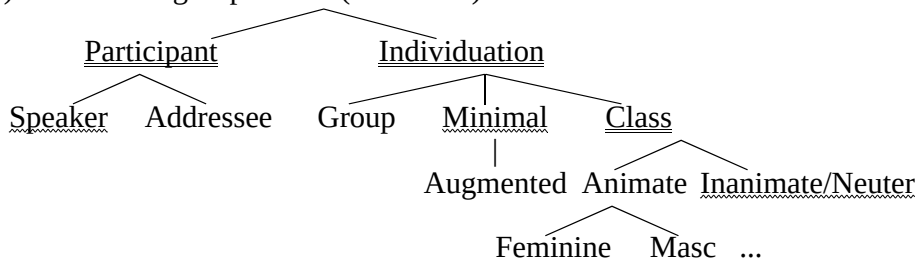
Markedness is encoded in a geometry via a node-counting metric. The more marked a given feature combination is, the more nodes will be required to represent it. In the tree in 6, it

takes more nodes to represent feature [G] than to represent simply feature [F]; therefore, a tree that includes feature [G] will be more marked than a tree that includes simply feature [F]. For further discussion, see Harley (1994).

Finally, in phonological feature geometries like those of Archangeli (1988) or Avery and Rice (1989), organizing nodes with no dependents receive a default interpretation, usually treated as underspecification: one of the daughter nodes is identified as representing the default interpretation of a bare organizing node. We make use of this notion in our discussion of the acquisition of pronominal paradigms below. In the tree representing our proposed geometry in 6 below, the underlined daughter nodes, Speaker, Minimal and Inanimate/Neuter, represent the default interpretation of an unmarked organizing node.

2. A morphosyntactic feature geometry. The geometry which we propose is provided by Universal Grammar is represented in 6. In any given language a subset of the possible features will be active — most languages will only use a portion of the features available.

(6) Referring Expression (=Pronoun)



In this geometry, all nominal features are dependent upon a root node which we call Referring Expression. We divide these features into three groups, identified by the nodes in small caps in the geometry. The Participant node and its dependents, Speaker and Addressee, will be used to represent person, specifically, 1st and 2nd person (3rd person being unmarked). The Individuation node and its dependents, Group, Minimal and Augmented, are used to represent

number systems. Finally, the Class node encodes gender and other class information. These three groupings represent explicitly via dependency the subgroupings of person, number and gender that other theories either assume without comment or represent via attribute-value matrices. This, then, is one way in which a feature-geometric representation accomplishes goal 4a for morphosyntactic features, that of defining natural classes of features for the application of morphological rules. In the remainder of this paper, we focus on how languages use the Participant and Individuation nodes to represent person and number and the interactions between them. We do not address the content of the Class node here, but see section 6.3.1 for discussion of the dependency relation between number and gender.

2.1 Two major classificatory nodes. It has long been recognized that there is a fundamental difference between 1st and 2nd person on the one hand, and 3rd person on the other. Two independent expressions of this insight are seen below:

> '"Person" belongs only to I/you, and is lacking in he.'
>
> (Benveniste 1971:217)
>
> 'Whoever does not act a rôle in the conversation either as speaker or as addressed remains in the great pool of the impersonal, referred to as "third person."'
>
> (Forchheimer 1953:5-6)

Forchheimer (1953) identifies a variety of morphological generalizations which show that languages treat 3rd person differently from 1st and 2nd, and, moreover, indicate that 3rd person is unmarked, relative to the other persons. His generalizations are listed in 7.

(7) a. 3rd person agreement is often zero, 1st/2nd person agreement is overt.

b. Many languages have no 3rd person pronoun — or at least no nominative form.

c. Many languages have distinct 1st & 2nd person pronouns only; for 3rd person they use demonstratives.

d. Closely related languages often have cognate 1st and 2nd person pronouns but 3rd person pronouns which are not obviously related.

e. 1st and 2nd person are often similar in form and inflection but dissimilar from that of 3rd person.

f. 3rd person is much more subject to objective subdivisions such as class, gender, and location.

(Forchheimer 1953:6)

We attribute this distinction between 1st and 2nd person, on the one hand, and 3rd person on the other, to the fact that the reference of the former is determined by the changing discourse roles, whereas the reference of the latter is fixed (cf. Jakobson 1971:131-132).[6] This is illustrated in the simple conversation in 8, where the referent of the pronoun I is person A for the first utterance and person B for the second. Similarly, the referent of the pronoun you is person B for the first utterance and person A for the second. In contrast, the referent of he remains constant across both utterances.

(8) a. A: I_A think he_C wants $your_B$ advice.

b. B: I_B think you_A're nuts. He_C doesn't want anything.

The discourse dependence of 1st and 2nd person pronouns also shows up in binding possibilities for pronouns. In 9 and 10, substitution of a 1st and 2nd person personal pronoun for a reflexive direct object results in a marginally grammatical sentence with the same

interpretation as the grammatical sentence with the reflexive. On the other hand, in 11, substitution of a 3rd person personal pronoun for a reflexive direct object results in a completely ungrammatical sentence if the subject and object are coreferential, and a completely grammatical one if the subject and object have disjoint reference:

(9) a. I love myself

b. ??I love me.

(10) a. We love ourselves.

b. ??We love us.

(11) a. He loves himself.

b. *He_i loves him_i.

c. He_i loves him_j.

The conceptual distinction between discourse dependence and independence is the external factor which determines the shape of the geometry in 6. The Individuation node represents those features of a DP which are independent of the discourse, that is, its number and gender. In contrast, the Participant node and its dependents represent those features which depend on the DP's discourse role.[7]

The geometry in 6 also captures the intuition that so-called 3rd person is in fact not a true personal form. As Benveniste emphasizes, 'the ordinary definition of the personal pronouns as containing the three terms, I, you and he simply destroys the notion of "person"' (1971:219). This intuition is encoded in our geometry because the Participant node represents only 1st and 2nd person. When the Participant node is absent, the underspecified Referring Expression node receives a so-called 3rd person interpretation.

We include suggestive paradigms from Maltese, a Semitic language, and Lyélé, a Niger Congo language of Burkina Faso in tables 1 and 2, which highlight the split between discourse participants and others. (Here, and throughout the paper, we represent the forms as they appear in our original source document.) In Maltese, gender distinctions are manifested in 3rd person only, as is the case in English and other familiar languages.

INSERT TABLE 1 ABOUT HERE

The Lyélé paradigm in table 2 provides even more striking evidence for this split in that formality, a discourse dependent property, is only manifested in 1st and 2nd person forms.[8] And as in Maltese, only 3rd person forms show marking for discourse independent properties, such as animacy, humanness, size and the count/mass distinction.

INSERT TABLE 2 ABOUT HERE

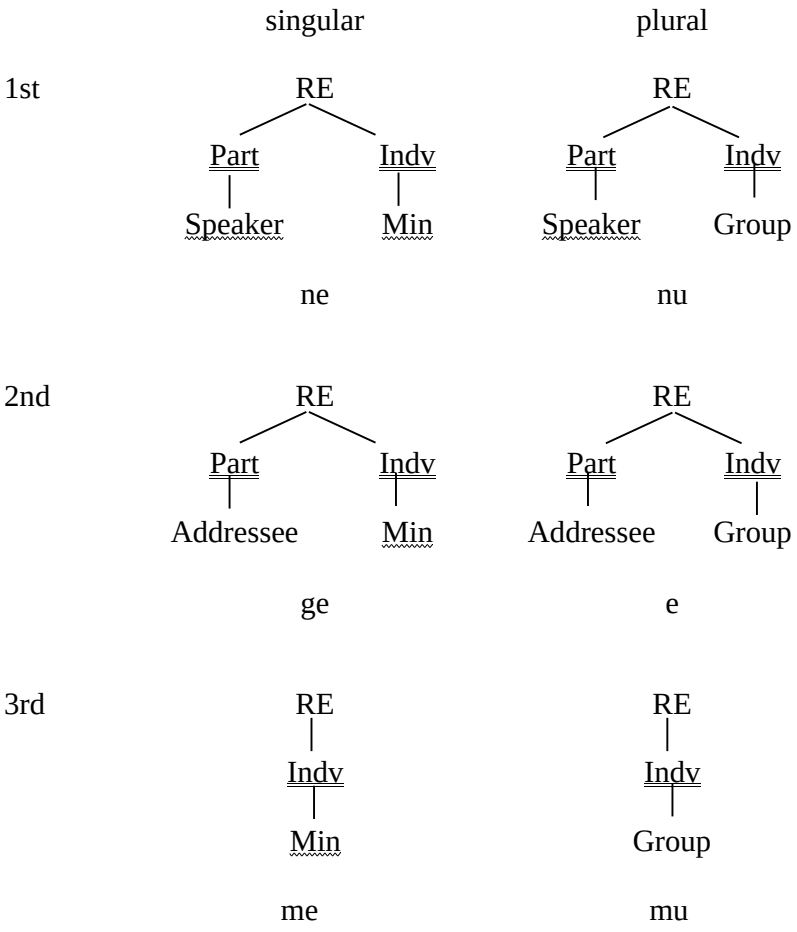
2.2 Illustrating the geometry. Before going on to discuss some paradigms with particularly interesting properties for the system we have so far proposed, let us introduce the geometry piece by piece, ultimately illustrating in detail a language which exploits the full range of possibilities allowed by the person/number structure.

2.3 Daga. First, let us consider a representative language with a minimum number of distinctions, regularly made. Daga, a language of the Trans-New Guinea family, fills all the paradigm spaces for three persons and two numbers in its emphatic pronouns (Murane 1974).[9] Its pronominal paradigm can be seen in table 3 below:

INSERT TABLE 3 ABOUT HERE

Languages like Daga, which distinguish 3 persons and 2 numbers, use the Participant, Addressee, Individuation and Group nodes of the geometry, as illustrated below:

(12) Geometries for Daga pronouns:



Person distinctions are made under the Participant (Part) node, number distinctions under the Individuation (Indv) node. We have represented the underspecified values for Participant and Individuation here explicitly, underlining them to indicate their default status. Speaker (1st person) is the default person; Minimal (singular) the default number. These nodes need not be explicitly represented in the geometry for 1st person singular, for example, but may be filled in

by default rules (see discussion in section 3.1 below). The 1st person singular pronoun, therefore, is less marked than the 1st person plural, despite the fact that the geometries for those two pronouns above appear to contain the same number of nodes: the node-counting metric for markedness applies to underspecified representations, without the default features. We must have the nodes for the default features available in the geometry, however, as they play a contrastive role in many languages when activated. Note that third person forms will be the least marked, as they do not require the presence of the Participant node.

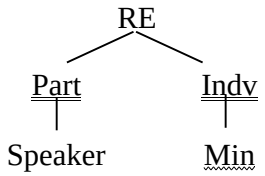
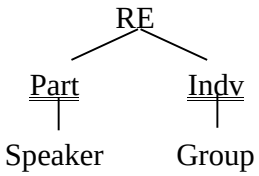
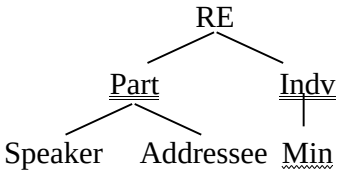
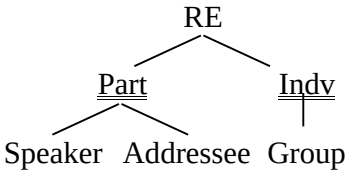
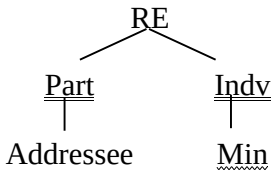
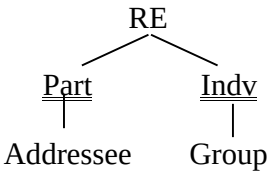
Number distinctions are represented by the features which depend on the Individuation node. Intuitively speaking, Individuation sorts entities in the world according to their discourse-independent properties, that is their quantity and class. Singular entities are represented by the Minimal node, which, like Speaker, is usually a default value for an underspecified organizing node. Group identifies multiple entities, and its presence represents plural number.[10]

2.4 Kalihna. Next, we consider the geometries active in a language that represents four persons and two numbers; our representative language is Kalihna, a Carib language spoken in French Guiana, Guyana, Surinam and Venezuela (Hoff 1968), which has a four-person, two-number paradigm of emphatic pronouns. Its paradigm is represented in table 4.

INSERT TABLE 4 ABOUT HERE

The geometries that represent these pronouns are illustrated in 13 below.

(13) Kalihna geometries

	singular	plural
1st excl.	 au	 aʔna
1st incl.	 kïxko	 kïxka:ro
2nd	 amo:ro	 amïiyaro
3rd	 moxko	 moxka:ro

The significant difference between the geometric possibilities in this language and those in Daga is the contrastive, fully specified availability of the Speaker node (hence it is not underlined in the above representations). Like many other theorists, we treat ‘inclusive’ forms as the conjunction of a Speaker and an Addressee feature (see, for example, Dalrymple and Kaplan 2000:784). The 1st person inclusive forms above have a maximally complex Participant node.

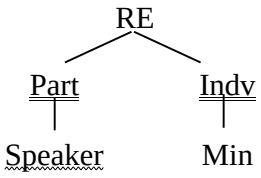
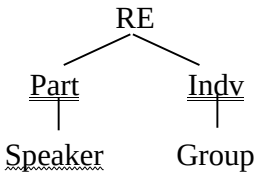
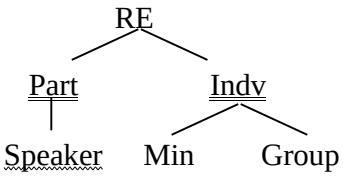
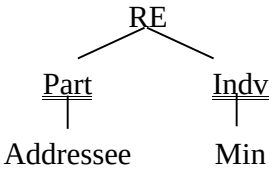
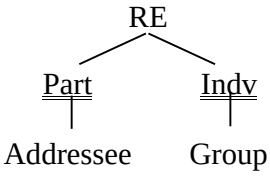
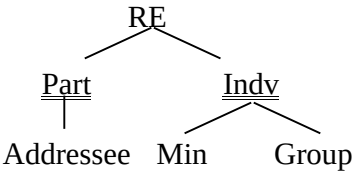
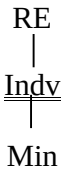
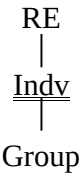
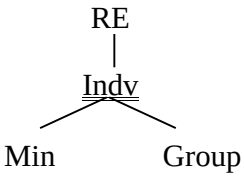
2nd person forms are represented by the projection of the Addressee feature without the Speaker node. 1st person exclusive forms are represented by the projection of the feature Speaker without the Addressee node. The learner can deduce that Speaker is not underspecified in her language from the presence of this inclusive/exclusive contrast. In this way, we capture the fact that inclusives are more marked than other persons, cross-linguistically, and yet allow for languages which make reference to both Speaker and Addressee features in their morphophonological forms for inclusives. Theories which represent inclusives with an [+incl] feature are unable to characterize such forms naturally (see the discussion of Ojibwa in section 4.2 below).

2.5 Tonkawa. For a representative language making three person and three number distinctions, we turn to Tonkawa, a Coahuiltecan language described in Hoijer (1933). Its nominative pronominal paradigm is in table 5 below:

INSERT TABLE 5 ABOUT HERE

The difference between two numbers and three is accomplished by allowing a fully specified, contrastive Minimal feature as a possibility in the geometry, parallel to the use of the Speaker node to move from three persons to four above. The geometries which correspond to these pronouns are illustrated in 14:

(14) Tonkawa geometries

	singular	plural	dual
1st			
	ca'ya	geuca'ya	geuca'ga
2nd			
	na'ya	wena'ya	wena'ga
3rd			
	'aye'la	'awe'la	'awe'ga

The addition of dual to a number system is attributed to the simultaneous activation of Minimal and Group, just as inclusive was represented by Speaker with Addressee above. The combination of Minimal and Group captures the intuition that the smallest possible non-singleton set contains two entities.[11]

Evidence for this somewhat unorthodox approach to dual number comes from the way in which dual is marked in Hopi, an Uto-Aztecan language of Arizona. In Hopi, dual can be expressed as a combination of minimal (≈ singular) and group (≈ plural) marking on different parts of speech (Ken Hale p.c.).

(19) a. Singular

Pam taaqa wari

that man ran.sg

'That man ran'

b. Plural

Puma taʔ-taq-t yuʔtu

those pl-man-grp ran.pl

'Those men ran.'

c. Dual

Puma taaqa-t wari

those man-grp ran.sg

'Those 2 men ran.'

When the subject is singular and verbal agreement is singular, singular number is conveyed. When the subject is plural and verbal agreement (suppletive in this case) is plural, plural number is conveyed. However, when the subject is plural and verbal agreement is singular, dual number is conveyed (see 19c).[12] Since in our system, dual consists of a co-occurrence of the usual plural and singular features, morphological reflexes of each are expected to surface in some cases. Note also that the shapes of the Tonkawa pronouns support this analysis: the dual appears to be made up a prefix from the plural attached to a singular base. See also the discussion of Koasati dual in note 38 below.

Our account of the dual predicts that no language will have a dual number unless it also has a plural number — that is, unless the Group node is independently active in the language. Feature bundle systems do not make this prediction unless it is externally stipulated.

2.6 Chinook. Chinook, a language of the Penutian family, makes four person and three number distinctions. Its paradigm, as described by Boas (1911b), is illustrated in table 6 below: [13]

INSERT TABLE 6 ABOUT HERE

The geometries for Chinook will represent a combination of the person geometries of Kalihna and the number geometries of Tonkawa, with the 1st person inclusive dual form having the most complex geometry, consisting of activated Speaker, Addressee, Minimal and Group nodes.

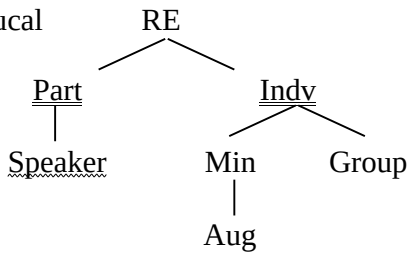
2.7 Yimas. Moving on to a language with three persons and four numbers, we turn to Yimas, a Sepik-Ramu language of Papua New Guinea, as discussed by Foley (1991). In addition to the dual, Yimas includes a fourth number, the paucal, representing a small group. Its paradigm is in table 7. The third person form included here is a masculine, proximal demonstrative, not a personal pronoun. We include it to fill out the paradigm on the assumption that the same representational constraints hold for the person and number features of demonstratives.

INSERT TABLE 7 ABOUT HERE

Yimas shows some interesting properties, not least of which is the syncretism between 1st and 2nd person paucal forms, and the absence of a paucal in the demonstrative system. We focus here on how to represent a paucal number at all. Recall from the discussion above that a language never has a dual without also having a plural (captured by the fact that both Group and Minimal must be active to encode dual). Similarly, no language has a paucal without also having

a dual.[14] In geometric terms, this entails that the paucal must be represented by the addition of another node (making it more marked) and that that node is dependent on the node that permits the expression of dual — that is, the contrastive Minimal node. We propose a feature Augmented, to capture the notion that, conceptually, the paucal consists of the smallest possible group (two) plus one (trial) or a few more (paucal).[15] This gives a representation for the 1st person paucal form illustrated in 15 below:

(15) 1st paucal

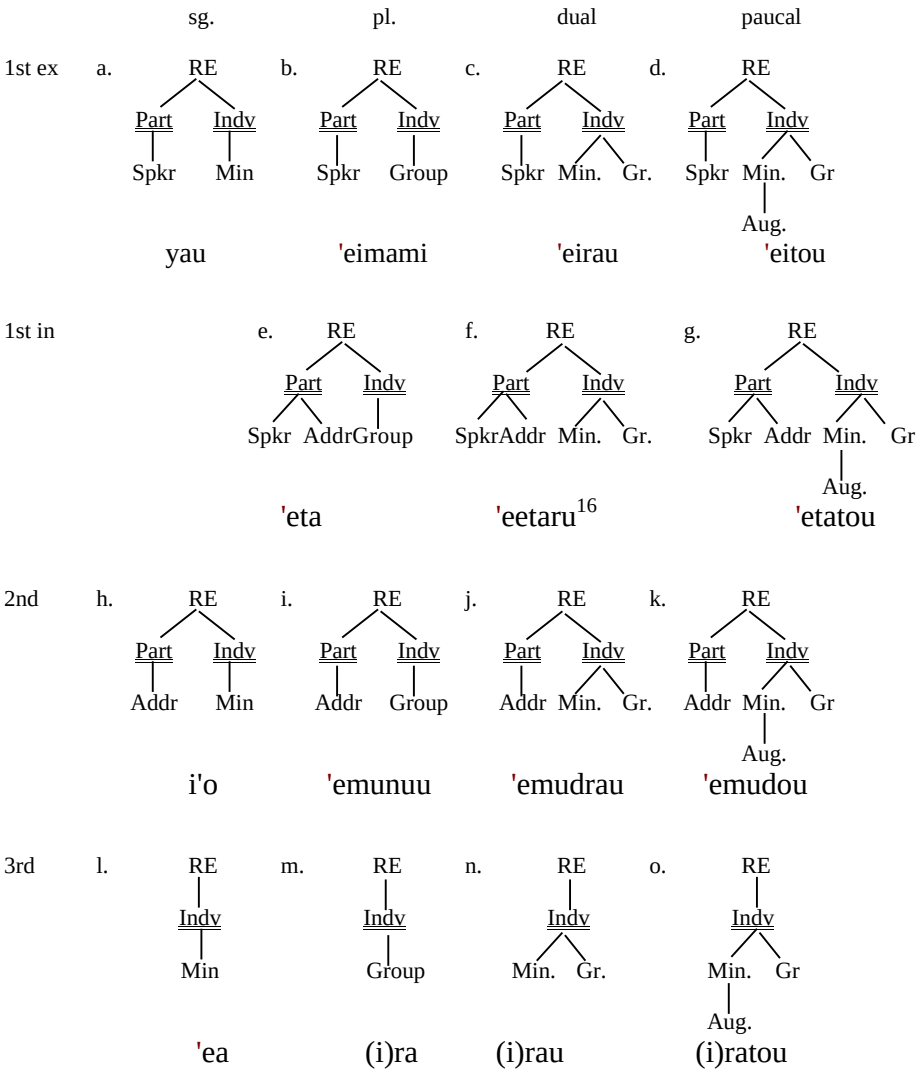


2.8 Boumaa Fijian. Finally, we turn to a language which exploits the full range of person/number possibilities. Table 8 shows the paradigm for the cardinal pronouns of Boumaa Fijian:

INSERT TABLE 8 ABOUT HERE

Note especially that Boumaa Fijian makes an inclusive/exclusive distinction, and exhibits four numbers: a singular, a plural, a dual, and a paucal (a small group). In 16, we illustrate each pronominal form with its associated geometry. Notice that the most geometrically complex, and hence most marked form, intra- and cross-linguistically, is the 1st inclusive paucal pronoun, which uses all eight nodes of the geometry.

(16) Boumaa Fijian geometries



No language has more than four numbers, and so the geometry permits all and exactly the attested person and number distinctions in the world's languages.

From the above, it should be clear how languages with straightforward pronominal paradigms distinguishing from six to fifteen person/number combinations will represent each pronoun geometrically.

2.9 Typology of person and number in pronominal systems. We close this section by noting that the relative complexity of the different kinds of pronominal systems illustrated above corresponds roughly to their frequency cross-linguistically, as documented in a pronominal database of one hundred ten languages representing over ninety distinct families or subfamilies, four isolates and three creoles.[17] In compiling the figures below, languages from the same subfamily were grouped together and counted as one; so, for instance, the Germanic languages of Indo-European count as one language, the Romance languages of Indo-European as another.[18] See the full listing in the Appendix, broken down by system complexity, for details.

First, breaking down languages by person system and number system, we see that frequency correlates fairly well with the predictions of the geometry, in tables 9 and 10. [19] Systems with inclusive persons are rarer than systems without; systems with dual number are rarer than systems without, and systems with paucal number are very rare indeed.

INSERT TABLE 9 ABOUT HERE

INSERT TABLE 10 ABOUT HERE

Setting aside for the moment the question of numberless systems (see section 4), let us consider the results for the geometry of the intersection of person and number, as outlined above; those results are summarized in table 11. By far the most common pattern is three persons and two numbers (forty-five languages); the next most common is four persons and two numbers (twenty-five languages), then three persons and three numbers and four persons and three numbers (eleven and seven languages, respectively), followed by the extremely rare three persons/four numbers and four persons/four numbers (one and two languages, respectively).

INSERT TABLE 11 ABOUT HERE

In general, then, the overall pattern evident in frequency of occurrence is consistent with the theory of feature representation presented here.[20]

2.10 Predictions, accounting for Greenbergian universals. The treatment above makes strong predictions about co-occurrence of distinctions. They are enumerated in 17:

(17) a. A language will not have a dual number if it does not have a plural number.

b. A language will not have a paucal/trial number if it does not have a dual number.[21]

c. A language will not have an inclusive person if it does not have a 2nd person.

Depending on the specific inventory of features employed in a feature-bundle approach, some of these predictions are made by those approaches as well. For instance, the [Speaker] and [Hearer] features of Dalrymple and Kaplan (2000) can capture 17c. However, the predictions 17a,b are not captured by any approach that employs a [dual] feature and/or a [trial] feature as feature-bundle alternatives to or co-occuring with [pl]; their mutual dependence must be externally stipulated.

In section 6, we return to the question of universals, and discuss possible treatments of some problematic pronominal systems. We now move on to discuss the acquisition of pronouns, and how it motivates both the overall geometry and the underspecification for Speaker and Minimal alluded to above.

3. Acquisition evidence for a feature geometry. Our assumption that this geometry is provided by Universal Grammar (UG) makes strong predictions about the acquisition of personal pronouns. In this section we demonstrate that the morphological feature geometry allows us to predict the observed uniformity and variability in the order of emergence of personal pronouns observed in ten studies spanning six different languages. These facts not only provide independent support for this geometry, but also motivate universal defaults for major organizing nodes, specifically 1st person for Participant and 3rd person singular (inanimate) for Individuation. The analysis in

this section was developed in collaboration with Rebecca Hanson; see Hanson (2000) for additional discussion.

3.1 Acquisition of the feature geometry. The acquisition of pronouns, specifically, their order of emergence, has received only sporadic attention in the literature, and there has been no previous attempt to approach the process using a feature geometry. There is, however, a good deal of research in phonology dedicated to the acquisition of a feature geometry, focussing on the emergence of segments and contrasts. We draw especially on Rice and Avery (1995) and Brown (1997) in our approach to the acquisition of a morphological feature geometry.

Rice and Avery (1995) develop a model of phoneme acquisition that attempts to capture both the global uniformity and local variability in the acquisition of sounds. "Global uniformity" refers to the observation that all children ‘acquire roughly the same set of basic sounds in roughly the same order’ (Rice and Avery 1995:25). "Local variability" refers to the fact that different children will acquire sounds from this basic set in different orders. Assuming that UG includes a phonological feature geometry, they propose essentially that language acquisition is a structure building, rather than structure pruning, process. 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. In this way, the geometry captures the global uniformity apparent in child language. However, different organizing nodes may be elaborated in any order, thus allowing for local variability.

A further assumption of Rice and Avery (1995) and Brown (1997) is that redundant information is absent from underlying representations. Both works assume Minimal Contrastive Underspecification: if a node is filled-in by a default rule provided by UG, it does not appear in the underlying representation. We assume, however, that if a node has, or can have, dependents,

or if its presence is contrastive in the language, it must be represented in the underlying representation, and no default fill-in rules for that node apply.

In short, following Rice and Avery (1995) and Brown (1997), we make the assumptions about the acquisition predictions of feature geometries listed in 18:

(18) a. The Structure Building Hypothesis

UG provides a minimal initial structure, which is added to in response to contrasts detected in the input.

b. Constrained general learning path

Acquisition proceeds from the top down; a given node must be acquired before its dependents. In this way the geometry captures the global uniformity apparent in acquisition.

c. Free specific learning path

The available paths may be elaborated in any order. In this way the geometry captures the local variability in acquisition.

3.2 The acquisition of pronouns: Evidence for universal defaults. The studies from which we draw our data are summarized in table 12 below. In all cases, the first pronouns acquired as recorded in the study are placed in column 1, the second in column 2, etc. Note that entries in the same column of the table do not necessarily correspond to the same age or even the same stage of acquisition – instead, they reflect the relative order in which the different pronouns are acquired.

Table 12 is arranged in three blocks. Block I includes studies A to D, which recorded 1st person as the initial pronoun. Block II consists of rows E, F and G and includes the studies that recorded 3rd person as the initial pronoun. Finally, rows H, I and J constitute Block III. This block includes the studies that began at a point where more than one pronoun was already acquired; thus, it isn't clear which pronoun emerged first. Note, however, that in all the Block III studies, the children always had both 1st singular and 3rd neuter from the initial session.

INSERT TABLE 12 ABOUT HERE

A comparison of the observations reported in the different studies listed in table 12 reveals significant variability across children in the order of emergence of specific personal pronouns. Note that out of the ten studies considered, no two report the same order of acquisition. However, there are some discernible patterns in both the uniformity and the variation observed, which are summarized in 19.

(19) a. <u>uniformity</u>:
- (i) initially 1st person singular or 3rd singular neuter/inanimate
- (ii) 2nd person after 1st person
- (iii) singular before plural

b. <u>variability</u>:
- (i) 2nd person relative to 3rd person (animate)
- (ii) 2nd person relative to plurals

These patterns, particularly those in 19a, strongly suggest that there are default interpretations for each of the organizing nodes in the geometry. The consistency in the initial pronoun as either 1st singular or 3rd singular inanimate conforms to the generalization that defaults are acquired first. Overall, the acquisition data indicates that the defaults are the following:

(20) Proposed UG-supplied Defaults

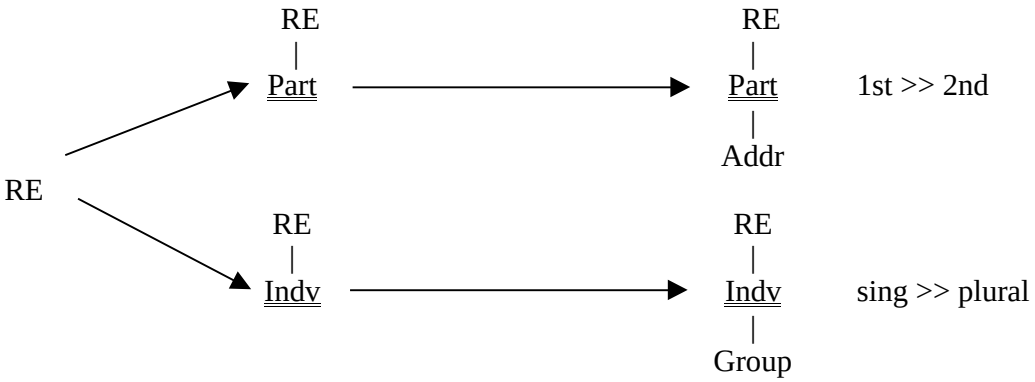
a. Speaker for Participant – therefore '1st person' acquired early

b. Minimal for Individuation – therefore 'singular' acquired early

c. Inanimate for Class – therefore 'neuter/inanimate' acquired early

A Speaker default at the Participant node is consistent with the early acquisition of 1st person; likewise, a Minimal default at the Individuation node allows us to predict that singular should also emerge early on. These two acquisition paths, illustrated schematically in 21 below, are predicted by the fact that Participant and Individuation have equal status in the geometry.

(21)



Because both Participant and Individuation are immediate dependents of the root RE node, 1st and 3rd person are equally likely to appear as the first pronoun. Likewise, since Addressee, Group and Class are all equally embedded in the geometry it is also predicted that they should be acquired later and that the relative order among them should vary, as seen in Table 12.

3.2 Against the acquisition of a feature bundle. It is worth noting that both the consistency and the variation in the data are problematic for a model that assumes that features are represented in unordered bundles. Compare, for example, the feature bundles in 22a and 22b, representing 1pl

and 2pl respectively. If the features for '1st person', '2nd person' and 'plural' are unordered, we would expect to see very little consistency in their relative order of acquisition. In particular, the feature bundle model does not make any predictions about why 1st person comes before 2nd person; a similar problem arises with the acquisition of number.

(22)

a. $\begin{bmatrix} +1 \\ -2 \\ pl \end{bmatrix}$ (1pl)

b. $\begin{bmatrix} -1 \\ +2 \\ pl \end{bmatrix}$ (2pl)

We might therefore be led to assume a hierarchy of person and/or number features to account for the consistent early emergence of 1st person and singular pronouns. As noted above, this kind of hierarchy has been independently proposed, and there are several versions of it. For illustration, consider the one proposed by Noyer (1992), discussed above in broad form in example 2 and in more specific form reproduced as 23 below:

(23) Noyer's hierarchy: 1 > 2 > <u>pl</u> > <u>dual</u> > <u>f</u>

This model is able to account for only some of the acquisition data. For example, the French study in Row C of table 12 conforms to its predictions. However, the Hebrew study in row G poses a problem because the feminine 3rd person pronoun was acquired before both 2nd person and plural.

Specifically to account for the acquisition data, Chiat (1986) proposed a different hierarchy, involving only person features, in which the order was 1st person >> 2nd person >> 3rd person. However, this model also proved inadequate. The most obvious difficulty, as Chiat noted, is the fact that the 3rd person inanimate pronoun consistently emerges alongside 1st person, rather than with the other 3rd person pronouns as expected.

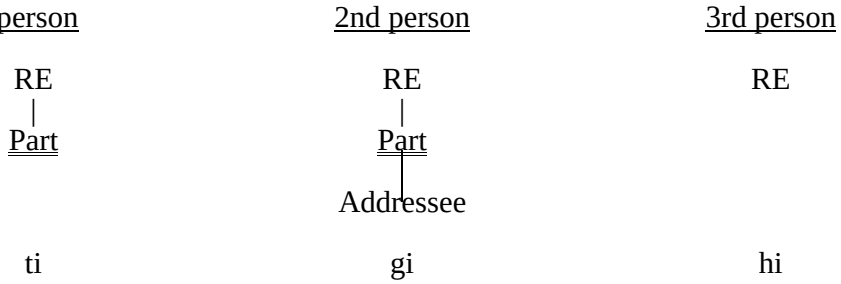
In summary, it is not clear how previous models of feature organization can account for the acquisition facts. A model of unordered feature bundles, with or without a supplementary hierarchy, has particular difficulty with the variation: if features are unordered, there should be more variation than there is; and if they are ordered in a simple hierarchy, there should be less. On the other hand, we have seen that the feature geometry, with the proposed defaults, correctly predicts where there should be uniformity and where there should be variability in the acquisition data.

4. Variability of interpretation: the effect of empty paradigm space. In theory, a language with an elaborate set of person distinctions, and no number distinctions whatsoever could exist, given that person, represented by activation of the Participant node and its dependants, and number, represented by activation of the Individuation node and its dependants, are structurally equivalent but distinct kinds of features.[22] Pirahã, an Amazonian language, seems to instantiate such a system. It has just three true personal pronouns, as shown in table 13.

INSERT TABLE 13 ABOUT HERE

This paradigm is represented in feature-geometric terms in 24. Notice, in particular, that it does not require an Individuation node at all.

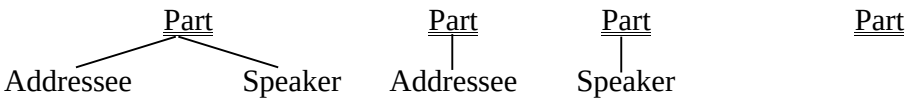
(24)



This does not exhaust the geometric possibilities available to languages that lack an Individuation node, however. Restricting our attention to representations activating all or part of

the Participant subgeometry, there are four possibilities (including the unmarked Participant node, available when Speaker is contrastive), as illustrated in 25:

(25) Person: 4 logical possibilities



In this section we discuss the analysis of languages that appear to make number distinctions, but only in the 1st person. We argue that these rare languages are in fact without 'true' number distinctions, and that their paradigms are constructed using only the four distinctions made available by the Participant node, illustrated above. Note that these are to be distinguished from languages that do not express number distinctions in pronouns but have such distinctions elsewhere, typically on common nouns or in verbal agreement. (See, for example, our approach to Koasati in section 6.1.) Our proposal involves co-opting the interpretation of the restricted available geometries to fill out the semantic paradigm space, within limits imposed by the nature of the features themselves. [23]

4.1 Four person geometries: Maxakalí and Kwakiutl. In previous work (Harley and Ritter to appear, Ritter and Harley 1998), we discussed the facts of Maxakalí and Kwakiutl, both of which have inclusive pronouns. These languages, whose paradigms are reproduced in tables 14 and 15 respectively, are notable in that they appear to manifest number distinctions in 1st person only. However, both languages have three 1st person pronouns: a singular, an exclusive plural and an inclusive.

INSERT TABLE 14 ABOUT HERE

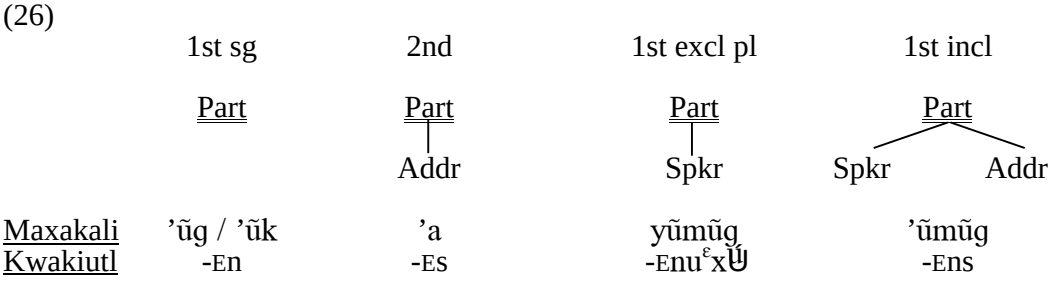
INSERT TABLE 15 ABOUT HERE

It is true of 1st person plural forms in general that they denote a mixed group consisting of the speaker and other individuals. This is in marked contrast to 2nd and 3rd person plural forms, which may denote a group of addressees or a group of other individuals, respectively. In natural language, there is no genuine 1st person plural—we never speak in choruses.

Exploiting the fact that 1st person plurals do not refer to a group of speakers, but rather to one speaker and either one or more addressees, or a speaker and one or more other individuals, we argue that the pronouns in these languages contrast only in person, and that in fact they do not express number at all. Support for this approach comes from the observation that these languages, like Pirahã, normally make no morphological number or gender distinctions on nouns or verbs.

We propose that Maxakalí and Kwakiutl make full use of the four different Participant sub-geometries available in the system. These sub-geometries are depicted in 26:

(26)



Maxakali	’ũg / ’ũk	’a	yũmũg	’ũmũg
Kwakiutl	-Ɛn	-Ɛs	-Ɛnuˁxʉ̋	-Ɛns

The fact that these languages make no number distinctions elsewhere in the grammar strongly suggests that number features are simply not present in their feature inventories. Thus, the only way to capture the contrasts among the four distinct 1st and 2nd person pronouns is by means of person features.

On this proposal, the geometries for first person singular, second person, and first person inclusive are entirely conventional. The activation of the Speaker node as contrastive (via the

evidence provided by the inclusive form), however, makes another potential geometry available to the language: Participant with a lone dependent Speaker node. These languages exploit that representation to add an additional contrast to their paradigm space: the 1st exclusive plural is represented with a dependent Speaker node, and is interpreted as referring to a group that includes the speaker — a first person plural. (Of course, in the absence of an Individuation node in these languages, 3rd person is represented as a bare RE node, as for Pirahã above.)

This analysis makes a prediction that no language that uses person distinctions as substitutes for number distinctions also has dual or trial/paucal numbers. This is certainly borne out in the languages discussed above, and in Aceh, the only other language of this type in our database (cf. Durie 1985). More generally, we predict that no language that uses person distinctions as substitutes for number distinctions will have more than four distinct 1st and/or 2nd person pronouns, as there are only four possible geometries available to a language using only the Participant node.[24] See section 6.1 for further discussion.

4.2 Over-riding the defaults: 2nd person inclusive. Finally, we turn to a much-discussed question: whether or not an inclusive pronoun is ever truly a 2nd person, rather than a 1st person form. That is, since both the Speaker and Addressee nodes are active in an inclusive form, one might expect that the morphological shape or other properties of the inclusive pronoun pattern with 2nd person, rather than 1st person forms. In our geometry, the Speaker and Addressee nodes are sisters, meaning that when they are both active as in an inclusive form, it is at least theoretically possible for either to be salient. Given our approach to universal defaults, according to which 1st person forms in general are less marked than 2nd, a 2nd inclusive would presumably be rare, requiring robust morphological evidence for acquisition. Nonetheless, it is a possibility in the system presented here, but not in a system which identifies the 2nd person

merely as [-Speaker, +Participant]. We argue that robust morphological evidence for a 2nd person inclusive exists in three distinct languages. Consider the paradigms in tables 16, 17 and 18, from Yokuts (Penutian), Ojibwa (Algonquian) and Nama (Koisan) [25]:

INSERT TABLE 16 ABOUT HERE
INSERT TABLE 17 ABOUT HERE
INSERT TABLE 18 ABOUT HERE

Note that in each case, the inclusive form patterns morphologically in the prefix with the 2nd person forms. Despite the shape of these pronouns, some theorists, including Zwicky (1977) and Noyer (1997), have argued that this similarity is not evidence that the inclusive form is 2nd person. However, Déchaine (1999) gives two convincing syntactic arguments that, in fact, the inclusive forms for Ojibwa in table 17 should be analyzed as 2nd person. First, the inclusive form may be used in the imperative, which is otherwise restricted to 2nd person forms. Second, there are two types of argument agreement on the Ojibwa verb, direct and inverse. The direct form occurs when the subject is 2nd person and the object is 1st person, and the inverse form occurs when the grammatical relations of these persons are reversed. The pattern is laid out in table 19. Crucially, the inclusive pronoun triggers the same agreement pattern as the 2nd person (exclusive), rather than the 1st person. That is, when the subject is an inclusive form, the marking is direct, and when the inclusive is the object, the inverse form occurs.[26]

INSERT TABLE 19 ABOUT HERE

In at least these two respects, then, it is the Addressee node, and not the Speaker node, that is the component of the pronoun crucial to the morphosyntax. We consider that these facts constitute robust evidence for a separate Addressee feature.

5. Delimiting the applicability of the geometry. It is obvious that there are many morphosyntactic features and feature distinctions that we have not touched on in the above

discussion, some of which bear directly on the predictions made by a feature-geometric approach, others which we have excluded from consideration because we feel there are principled reasons to set them aside temporarily. In this section we outline the reasons why we set aside some putative counterexamples to the geometry and analyses we propose, and discuss some phenomena that we have excluded from the geometry.

5.1 Constraining the feature content of a single grammatical morpheme. First, it is worth noting that the geometry is designed to represent the feature composition of a single pronoun morpheme; polymorphemic pronoun words, which are produced by the juxtaposition of two or more feature bundles, may allow for combinations of features that might otherwise be unexpected. For example, in Spanish, as seen in table 20, the nominative and prepositional case forms of 1st and 2nd plural pronouns are marked for gender, while all other 1st and 2nd person pronouns are unspecified for gender. This constitutes a violation of Greenberg's (1963:96) Universal 45: 'If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also'.[27]

INSERT TABLE 20 ABOUT HERE

On closer inspection we find that, unlike other members of the paradigm, these pronouns are bi-morphemic, nos+otros/otras and vos+otros/otras 'we+others(m/f)' and 'you+others(m/f)' respectively. Thus, these pronouns consist of one morpheme expressing person and number, but not gender (nos or vos), and a second morpheme expressing number and gender, but not person (otros or otras). It is the combination of these two morphemes which gives rise to this unexpected feature specification.[28, 29] Crucially, these Spanish pronouns do not contain a morpheme whose feature content consists of 1st or 2nd person and gender, but not number. In section 6 below we consider languages which are more problematic in this respect.

5.2 Restriction to closed-class, context-free features. A second restriction in the applicability of this geometry is that it is designed to represent the distinctions in functional or grammatical morphemes, specifically pronouns. What distinguishes pronouns from substantive nouns is that their shared semantic content can be reduced to a closed set of features, including person, number, and in some languages, gender. For languages such as Thai or Vietnamese, which may have both an open class of noun substitutes and a closed class of pronouns, we restrict our attention to the latter. Noun substitutes are typically kinship terms or titles, and in some cases they function as nouns. For example, tua, a Thai 2nd person pronoun used to address an intimate also functions as a noun meaning 'body' (Uri Tadmor, p.c.). Thus, unlike pronouns, noun substitutes have morpho-semantic content which is primarily lexical in nature. In fact, in Thai, interpersonal relationships, rather than person or number, seem to be the major consideration in choice among competing forms (cf. Cooke (1968) and Campbell (1969) for details).[30]

Similarly, it has been observed that in particular discourse situations, speakers occasionally use pronouns whose grammatical person, number and or gender specification is incompatible with the properties of the referent. Some English examples are provided in 27 below.

(27) [Individual speaking to a close friend/spouse who is clearly not in a good mood]

a. Uh-oh, she's/we're in a good mood!

Cf. Uh-oh, you're in a good mood!

[Nurse talking to patient]

b. How are we feeling today, Mr. Smith?

Cf. How are you feeling today, Mr. Smith?

In such situations, power or solidarity considerations override grammatical ones. For example, in 27a, the speaker may be distancing him/herself from the situation, and in 27b, the speaker may be expressing solidarity with the addressee, cf. Brown and Levinson (1987). What is important to note here is that these atypical uses trigger the same formal agreement on the predicate as does the canonical use of these pronouns.[31]

This geometry is not designed to capture formality distinctions, which are determined by the social context, or case distinctions, which are determined by the syntactic context, despite the fact that both form part of the pronominal paradigms of the world's languages. Formality distinctions, like the discourse-licensed atypical usages described above, express such things as relative status of the speaker and addressee or other, and the speaker's attitude to the addressee or other. In languages with formality distinctions in the pronominal system, these factors — part of the social context — are normally taken into consideration in determining the choice of pronoun. Of course, ultimately, one would want to include at least case features in the geometry, since they are obviously grammatically relevant properties of the pronoun, but they will no doubt require the addition of another organizing node, for a couple of reasons. First, case differs from person, number and gender in that it is not reflected in external (verb) agreement, cf. Lehmann (1988).[32] Second, Greenberg's Universal 39, reproduced in 28, suggests that case is not necessarily in the same natural class as the other features considered here.

(28) Universal 39: Where morphemes of both number and case are present and both follow or both precede the noun base, the expression of number always comes between the noun base and the expression of case.

(Greenberg 1963:95)

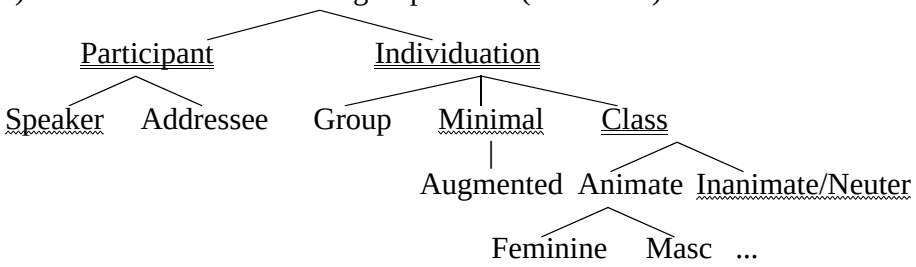
If case is either not expressed in this geometry, or if it is expressed by means of a separate node inserted or constructed by the syntax, we can account for its separation from the

morphological realization of these other features.[33] Finally, given that we assume that the nodes in the geometry are constrained and motivated by conceptual considerations, an investigation of the conceptual foundations of case systems would be necessary to incorporate them into the geometry. Such an investigation is beyond the scope of this paper.[34]

6 Predictions and problems. In this section we explore the limitations of the proposed geometry, identifying theoretically possible and impossible combinations of person, number and gender features. We then look at some actual pronoun systems that appear to defy our predictions, and explain why they are in fact not as problematic as they first appear.

6.1 Complex pronouns imply simpler ones. The feature geometry, reproduced here for convenience as 29, appears to permit almost any combination of person, number and gender features. For example, 1st or 2nd person features should combine freely with any of the number and gender features, since the latter are dependents of a separate organizing node.

(29)

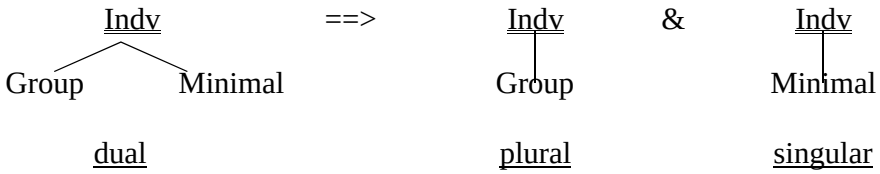


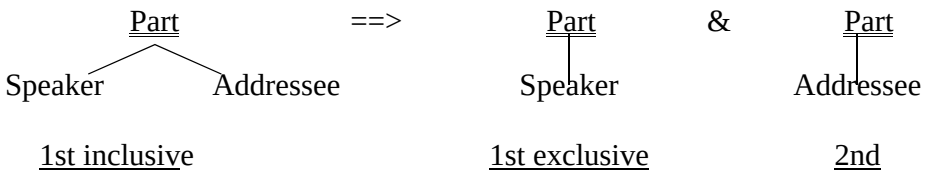
Assuming a model of feature acquisition along the lines of Rice and Avery (1995) or Brown (1997), we expect that if a language has a pronoun with a complex geometry, the simpler geometries that form the sub-constituents of the complex geometry are also available in the language, as illustrated schematically in 30:

(30) 

This constraint appears to shed light on a number of cross-linguistic generalizations regarding pronoun inventories. For example, dual never occurs unless the language also has singular and plural numbers, because dual requires the activation of both Minimal and Group nodes , and separately these nodes are used to express singular and plural, respectively. Similarly, inclusive pronouns only occur in addition to 1st person (exclusive) and 2nd person pronouns, because the inclusive person requires the activation of both Speaker and Addressee nodes, and separately these nodes are used to express 1st and 2nd person pronouns, respectively. These implicational relationships are illustrated in 31.

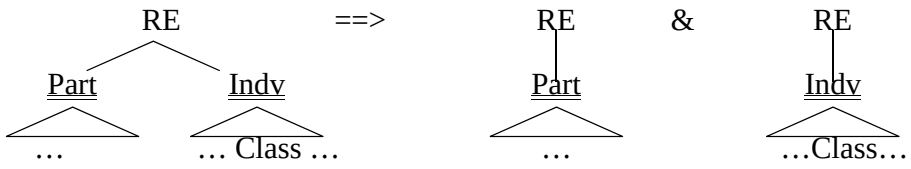
(31)

a. 

b. 

Extending this generalization would also allow us to account for the observation that only a minority of languages have any gender distinctions in 1st or 2nd person pronouns, and when they do, the 1st and 2nd person pronouns never have any gender distinctions that are not also found in 3rd person pronouns. This is due to the fact that a gendered 1st or 2nd person pronoun involves elaboration of both major organizing nodes but a gendered 3rd person pronoun only involves elaboration of the Individuation node.[35]

(32)

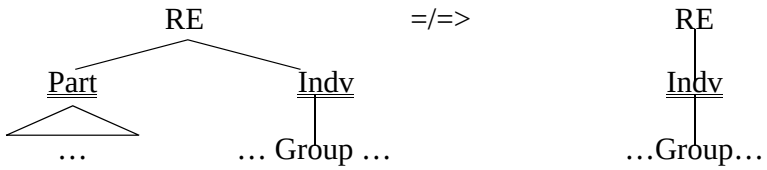


If the generalization holds for Participant and Class nodes, it ought also to hold for the Participant and Individuation nodes: there should never be more numbers in 1st and 2nd person than in 3rd. However, this implication fails for Guarani (Tupi) and Koasati (Muskogean), which have separate singular and plural pronouns for 1st and 2nd person, but only one 3rd person pronoun, unspecified for number (and gender).[36] For purposes of illustration, the Koasati paradigm is given in table 21 below; the Guarani paradigm has a similar structure.

INSERT TABLE 21 ABOUT HERE

For these languages, it appears as though the Individuation node has dependent nodes marking number, but only in the 1st and 2nd person — i.e. when the Participant node is activated.[37]

(33)



Given an interpretive theory of morphology like that of Anderson (1992) or Halle and Marantz (1993), where morphemes interpret the morphosyntactic features provided by the syntax, rather than contributing them, an account of this type of paradigm is possible in terms of syncretism: the different geometries are in fact present in the 3rd person, but the morphological inventory of the language does not provide distinct forms to realize them. For Koasati, at least, this appears to be the correct approach. Although number on 3rd person pronouns is not marked, number agreement with 3rd person nominal or pronominal arguments is remarkably robust in the

language — in fact, it is quite complex (Kimball 1991). In 53 below, we illustrate one subcase of number marking in Koasati; we assume that the presence of number agreement with 3rd person arguments signifies the activation of the Individuation node generally.

(34) Koasati singular, dual, plural

a. okipófka-k o:w-á:y
whale-subj loc-go.about.sg/du
'A whale is swimming about'

b. okipófka-k o:w-a:yá-:c
whale-subj loc-go.about.sg/du-3non.sg[38]
'Two whales are swimming about'

c. okipófka-k o:-yomáhl
whale-subj loc-go.about.pl
'Some whales are swimming about'

(Kimball 1991:446)

In Guarani, however, there is no obvious evidence that this is the right way to proceed; there is no number marking on nouns and number agreement on verbs is restricted to the 1st and 2nd persons. In the context of a plural 3rd person subject, an independent word, hikwaÊ is used, but only if there is no other 'word indicating plurality in construction with the verbal phrases' (Gregores and Suárez 1967:155).

(35) pe mita h-aseÃmá hikwaÊ

that boy he-cry-all plural

'Those boys are all crying'

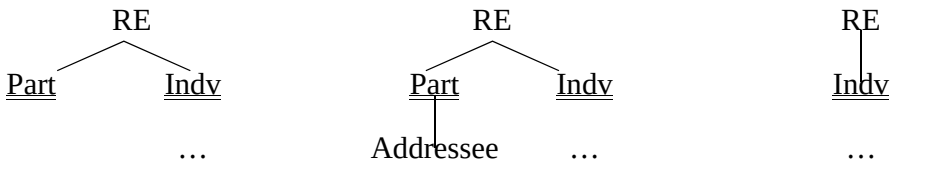
(Gregores and Suárez 1967:155)

The singular/plural distinction in both 1st and 2nd person pronouns confirms that the Group dependent of the Individuation must be active in this portion of the paradigm. As for 3rd person, it is unclear whether the independent plural word is a form of agreement, licensed by an active Group node in the representation of the 3rd person subject. If it turns out that hikwaÊ is in fact an agreement morpheme, then the account proposed above for Koasati can be extended to Guarani.[39]

6.2 The lower bound on geometric complexity. A further prediction of this framework is that we might find a system that is relatively less complex than those we have discussed to this point. In particular, we might expect to find a language that failed to activate one of the major organizing nodes. In section 4 we proposed that Pirahã, Maxakalí and Kwakiutl are examples of languages which do not have an activated Individuation node. It is known that many languages lack gender distinctions, including Acholi (Nilo Saharan), Brahui (Northern Dravidian), and Tonkawa (Coahuiltecan); we interpret this straightforwardly as a failure to activate the Class node.

Greenberg's (1963: 96) Universal 42 states that 'all languages have pronominal categories involving at least three persons and two numbers.' With respect to person distinctions, this can be represented as the claim that every language has at least the following set of distinctions in its inventory:

(36)



1st person 2nd person 3rd person

However, without external stipulation and assuming the geometric analysis, we might expect to find languages that are exceptions to Greenberg's universal because they fail to activate the Participant node altogether, or they fail to activate the Addressee node dependent upon it. In other words, there might be a language that lacks 1st and 2nd person pronouns altogether, or has a single pronoun or set of pronouns for all discourse participants. We conjecture that a language with no 1st or 2nd person pronouns is a language with no personal pronouns. In fact, this has been suggested of Japanese, Thai and other Southeast Asian languages, which use titles, kinship terms, names, etc. in place of pronouns.[40] If this is indeed the case, then these languages make no use of person features. We would expect that such languages would also lack verb agreement for person (as well as number and gender). This is certainly true of Japanese and Thai.

Are there languages that have a single pronoun or set of pronouns, used to refer alternatively to the speaker or the hearer? Noyer (1992), citing Lipkind (1945) and Ken Hale (p.c.), claims that the Winnebago (Siouan) singular pronoun nee has exactly this range of interpretations.

(37)

'I' or 'you'	nee
'he/she'	÷ee

(Noyer 1992:151)

Such a language would have only two person geometries as follows:

(38)

1st/2nd person 3rd person

However, Noyer notes that verb agreement is used to disambiguate between 1st and 2nd person interpretations of nee. If this is the case, then this, like Koasati number, can be reduced to syncretism in a paradigm which contains distinct 1st and 2nd person representations.

Michael Cysouw (p.c.) suggests that a more common scenario is exemplified by Awa (East New Guinea Highlands).[41] As shown in table 22, the independent pronoun paradigm for this language includes distinct singular pronouns for 1st person and 2nd person, but only one plural pronoun which may be used for either person. Again, on an interpretive morphological model, this is essentially the same phenomenon as that discussed above for Koasati 3rd person number: person distinctions undergo syncretism in the plural. The form ite spells out any geometry with a Participant node and a Group node, and is not sensitive to the presence or absence of any features dependent on the Participant node.

INSERT TABLE 22 ABOUT HERE

With respect to number, Greenberg's Universal 42 can be converted into a claim that every language has at least the following set of distinctions in its inventory.

(39)

singular plural

However, as we have already noted in our discussion of Kwakiutl, Maxakalí, and especially Pirahã above, there seem to be languages that fail to make any number distinctions,

although they do make a plethora of person distinctions. There are even languages, such as Pirahã that have a single pronoun for each person.[42] Given the absence of any grammaticalized number distinctions in the pronouns or elsewhere in the language (cf. Everett 1986), Pirahã can be treated as not activating the Individuation node at all. [43]

In short, Greenberg's claim that all languages have pronoun systems with at least three persons and two numbers is subject to exceptions, and the exceptions attested are in fact expected, given the geometry proposed here. What we predict not to exist are languages that use the same pronoun (or in a language with cases, the same set of pronouns) for both 1st and 3rd or both 2nd or 3rd persons. In fact, none of the languages we looked at has such a pronoun or set of pronouns in its inventory.[44]

6.3 Constraints on the expression of gender. Until now we have avoided discussion of questions of gender. We assume that gender is included in the feature composition of a pronoun when the representation includes an activated Class node. We assume that Class is a dependent of Individuation. However, we have not discussed the internal structure of this major organizing node. The reason for this is that gender (or class) features vary more widely in the world's languages than either person or number. For example, while all languages seem to have at most four persons and four numbers, the set of gender/class systems seems much less constrained. Some languages have no gender marking whatsoever, including Acholi, a Nilo Saharan language, Brahui, a Northern Dravidian language, and Tonkawa, a Coahuiltecan language. Other languages have two or three genders, and the limiting case is probably represented by the Bantu languages which have upwards of ten distinct genders or classes of nouns (cf. Corbett 1991). It may turn out that some systems involve an open-ended set of lexically determined classes while others involve a closed set of grammatically determined classes. The former would

of course be beyond the scope of our geometry. Consequently, we leave the problem of identifying the dependents of the Class node open for future research.

6.3.1 Dependency of gender on number. The reason that we assume that the Class node is a dependent of Individuation is that this captures Greenberg's (1963:95) Universal 36, which states that '[i]f a language has the category of gender, it always has the category of number.' To the best of our knowledge this universal is absolute.[45] At first glance, it is not obvious that our proposal does in fact capture this universal dependency of gender on number. In particular, how do we eliminate the possibility of a pronoun system that manifests gender distinctions, but not number distinctions? Such a system would require activation of the nodes Individuation, Class and at least one dependent of Class. However, it would crucially not activate either the Minimal or Group nodes, as schematized below:

(40)

default gender marked gender x

We suggest that the reason that such a system could not exist is that, as noted in section 6.1 above, complex geometries imply simpler ones. More specifically, a language with the geometries in 40 would also have a geometry consisting of a bare Individuation node in its inventory, and the latter would be uninterpretable in a system with gender, but not number.

6.3.2 Interactions of gender and number. The dependence of gender on number is also expressed in Greenberg's Universals 37 and 45, reproduced below:

(41) Universal 37: A language never has more gender categories in nonsingular numbers than in the singular.

(Greenberg 1963:95)

(42) Universal 45: 'If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also.'

(Greenberg 1963:96)

These observations are precisely what we expect to find given the geometric implications of complex representations. If a Class node may co-occur with Group, it follows that it may alternatively occur without Group, since a representation without Group is less marked.

Most commonly we find systems like those schematized in 43, where gender features are present in the singular, and perhaps also in the plural. The system in 43a is one where the singular pronouns are specified for gender (via activation of the Class node), and the plural pronouns are not. The former include a Class node dependent on Individuation but no Group node; the latter include a Group node dependent on Individuation, but no Class node. The system in 43b is one where both singular and plural pronouns are specified for gender (via activation of the Class node), and the plural pronoun requires activation of both the Group and Class nodes. Since this plural in 43b has an extra node compared to that in 43a, it is more complex and hence more marked, and we would expect this type of system to be relatively less common.

(43) Gender in singular, but not in plural

a.

b. Gender in singular and plural

Note that the representations in 43 are of course incomplete, as they lack the root Referring Expression, and in the case of 1st and 2nd person pronouns, the Participant node and its dependents. Once we incorporate these subgeometries into representations that also include the nodes specifying person features, the set of possibilities for a single language expands. In particular, a language might have both gendered and ungendered singular pronouns, and both gendered and ungendered plural ones, in different persons. In terms of our geometric representations, such a language would only have a Class node present in a subset of person-number combinations. And in fact, most commonly, we find gender distinctions restricted to 3rd person singular pronouns (e.g. Dieri (Australian) and Eastern Pomo (Hokan)), or to 3rd person singular and non-singular pronouns (e.g. Kaingáng (Macro-Gé) and Kannada (Dravidian)). Both kinds of systems are consistent with Universals 37 and 45.

Until recently it was thought that these universals were absolute. However, Plank and Schellinger (1997) demonstrate that there are a significant number of exceptions in a range of genetically distinct languages. Most of the exceptions they have found are marked, but not impossible systems in our framework. Some of the examples involve polymorphemic pronouns,

like the Spanish 1st and 2nd person forms treated in section 5.1 above. For example, the Windesi and Wandamen dialects of Central-Western Yapen (Austronesian) have two genders in the 3rd person plural, but not elsewhere (see table 23). However, as Plank and Schellinger (1997:63) point out, 'it is the suffix –at which distinguishes the human gender from the non-human one for 3rd person plural; and this same suffix is obligatorily present in 1st and 2nd person plurals.' In other words, human gender is expresed by a suffix on a pronominal stem which itself expresses person and number.

INSERT TABLE 23 ABOUT HERE

We suggest that exceptions of this type are amenable to the treatment outlined in section 5.1 above for Spanish, in which the polymorphemic pronouns are analysed as the result of the juxtaposition of two or more feature bundles. (See also the discussion of Lithuanian in note 28.)

Plank and Schellinger (1997:64) also discuss a type of exception, exemplified by Katu (Katuic, Mon-Khmer) and Palau (Austronesian) which, in their words, 'involves partial syncretism.' In these languages, a 3rd person singular animate or human pronoun is syncretic with an inanimate or non-human one. However, the latter is unspecified for number, leading to a paradigm such as that in table 24:

INSERT TABLE 24 ABOUT HERE

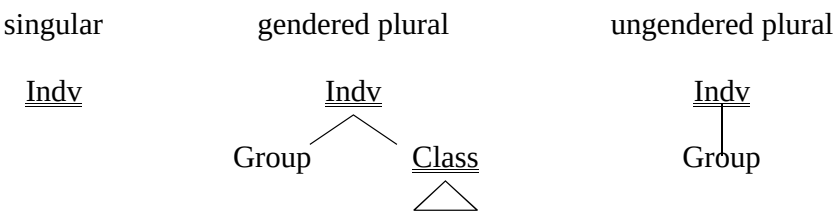
This class of exceptions is amenable to the treatment outlined in section 6.1 above for Koasati and Guarani. Given its L-shaped pattern of realization in the paradigm above, dó is clearly an elsewhere form, realizing a 3rd person geometry when no more fully specified form is available. We claim that Katu 3rd person forms do have distinct geometries for animate and inanimate in the singular, but that there is no pronoun in the inventory to realize the distinct 3rd person singular animate geometry, which is hence realized by the elsewhere form dó.

A third type of exception, exemplified by members of the Berber family, has gender distinctions in singular and plural numbers of some persons, and in plural number only of other persons. As Plank and Schellinger point out, this type of language would be permitted on a generous reading of Universals 37 and 45 which requires only that some singular pronoun be specified for gender. Such languages pose no problem for our model.

What we predict to be impossible are the systems schematized in 44 consisting of a more complex gendered plural together with a completely unmarked singular, realized as a bare Individuation node. Such a system contains singular pronouns which are only ungendered, and gendered pronouns which are only non-singular.

(44) a. Gender in plural, but not in singular

b. Gender in some plurals, but not in singular



Plank and Schellinger (1997) do list one language of this type: Biak, a language closely related to Wandamen and Windesi. The facts of this language are provided in 25:

INSERT TABLE 25 ABOUT HERE

The problem is that Biak appears to make an animate/inanimate distinction only in the third person plural, not in the singular, which has its own separate form. It is worth noting that

the gender distinctions appear only in the plural number, not in the dual or trial numbers. There are two different approaches that we could adopt here. On the one hand, we could assert that Biak is also exhibiting syncretism of underlyingly distinct geometries in the singular, and that i is a default, elsewhere form. This approach would need to be substantiated by independent evidence, perhaps from verb agreement, that the genders are indeed distinct in the singular. A different, somewhat more radical approach, would be to claim that Biak is a language which has overridden the UG-supplied default singular interpretation for the Individuation node. In Biak, the claim would be, the bare Individuation node would receive a plural interpretation; singular number would be represented by activation of the Minimal node, dependent on Individuation. In that situation, the least geometrically marked number would be plural, not singular, and therefore, gender distinctions should occur in the plural but not in the singular. Again, however, further investigation of the Biak number system would be required before we could substantiate this analysis. It would not be impossible for the child learning Biak to over-ride the defaults, as s/he would have positive evidence in the form of the plural gender distinctions; however, a more complex acquisition path would be expected. Further, we might expect to discover other evidence that plural was the unmarked number in the system; perhaps bare nominals in Biak receive a default plural interpretation.

The geometry that we have presented provides a representational explanation for the restricted combinations of person/number features attested in the pronouns of the world's languages. Depending on whether or not an interpretive morphological framework is adopted in conjunction with the geometry, certain exceptional patterns can be accounted for by assuming syncretism; others are the product of complexity. In terms of person and number features, as far as we know, no systems exist that are not predicted by the geometry. With a few exceptions,

discussed above, pronominal forms with gender also behave in accordance with the predictions of the geometry.

7 Conclusion. Feature-geometric organization of phonological features gave theoreticians a tool that captures the idea of a natural class of phonological features, permitting an explanation of certain feature groupings that arise in the phonological patterns of the world's languages. Morphological features group in the same way, although there are fewer of them, and the natural classes are not connected to their physical realization, as is the case for phonological features.

Here, we have proposed that a representational account of feature groupings extends naturally to morphological person and number features, and that the organization of the geometry is constrained by basic conceptual categories. We show how the proposal gives us insight into both the distribution of pronominal paradigms in the world's languages and the structure of unusual individual paradigms.

Like its phonological counterpart, the geometry proposed here makes predictions about possible and necessary contrasts within a given system, as well as predictions about what constitutes a natural grouping of features. These predictions are falsifiable, in two ways. A counter-analysis could present facts that the particular geometry proposed here cannot deal with, but which could be handled by a revised geometry, with different node values. Alternatively, it could be argued that the geometric approach itself is untenable, by showing that any geometry would be inadequate to account for a certain paradigm. For example, it might be the case that a form in a paradigm could show syncretisms which demonstrate that some node would need to be dominated by the Participant and Individuation nodes simultaneously (for instance, an inclusive that collapsed sometimes with 1st person singular forms and sometimes with 3rd person plural forms). If such a case could be found, it could demonstrate that the geometric approach to

morphological features is fundamentally flawed. It is our claim, of course, that no such case could exist.[46]

The idea of a morphological feature geometry as a representational tool in itself does not dictate a choice of theoretical framework; indeed, we feel that at least the concept of such a system of morphological features will be of interest to a broad range of researchers. In some of our specific proposals (e.g. with regards to Koasati), it may be apparent that we are thinking in terms of an interpretive morphology, like that of Anderson's A-Morphous Morphology or Halle and Marantz's Distributed Morphology; however, nothing about the concept of a feature geometry in itself requires such an approach. The insights the geometric approach provides are compatible with most extant morphological frameworks, although some of our specific proposals may not be.

Several major research questions now arise: the nature of the relationship between the geometry and the syntactic component, in particular with respect to agreement phenomena; the best way to integrate and represent gender and class features on the one hand, and case features on the other; the representation of verbal morphological features such as aspect, tense and mood and their interaction with the nominal feature system; and the spell-out of the geometry. We have demonstrated here that the geometric approach has enough intrinsic merit to warrant the directed investigation of these and related questions.

Appendix

a. Person: 1/2/3
Number: sg/pl 45 languages

LANG	Family	Subfamily	Notes
Acholi			
Ainu	Isolate		No 3rd person pronouns; demonstratives used for this purpose
Gilyak	Isolate		

Gulf Arabic, Hebrew	Afro-Asiatic	Semitic	
Hausa	Afro-Asiatic	Chadic	
Iraqw	Afro-Asiatic	Cushitic	
Central Tamazight	Afro-Asiatic	Berber	
Wolaytta	Afro-Asiatic	Omotic	
Bandjalang	Australian	Pama-Nyungan	
Basque	Basque		No 3rd person pronouns; demonstratives used for this purpose
Berik	Trans-New Guinea	Northern Trans-New Guinea	Sg/pl in 1st person only, no number distinction in 2nd, 3rd, but number and gender on verbs
Brahui	Trans-New Guinea	Northern Trans-New Guinea	
Daga	Trans-New Guinea	Main Section	
Tauya	Trans-New Guinea	Madang-Adelbert Range	
Cahuilla, Luiseño	Uto-Aztecan	Northern Uto-Aztecan	
Nahuatl	Uto-Aztecan	Southern Uto-Aztecan	
Albanian	Indo-European	Albanian	
Latvian	Indo-European	Baltic	
Dutch, English, German, Swedish	Indo-European	Germanic	
Greek	Indo-European	Greek	
Balochi	Indo-European	Indo-Iranian	No 3rd person pronouns; demonstratives used for this purpose
Catalan, French, Romanian, Spanish	Indo-European	Italic	
Polish, Serbo-Croatian	Indo-European	Slavic	
Welsh	Indo-European	Celtic	
Chechen	North Caucasian	N. Central Caucasian	
Kabardian	North Causasian	Northwest Caucasian	
Georgian	South Caucasian	Georgian	
Finnish	Uralic	Finno-Ugric	
Godié, Kongo,	Niger-Congo	Atlantic-Congo	

San Salvador, Southern Sotho, Swahili			
Haitian Creole	Creole		No number distinction in 2nd person
Nigerian Pidgin	Creole		
Halkomelem, Southern Puget Sound	Salishan	Central Salish	
Lillooet	Salishan	Interior Salish	
Kaingáng, Xokleng	Macro-Gé	Ge-Kaingang	
Rikbatsa	Macro-Gé	Rikbaktsa	
Kannada	Dravidian	Southern Dravidian	
Koasati	Muskogean	Eastern Muskogean	3rd person unmarked for number
Ladakhi	Sino-Tibetan	Tibeto-Burman	
Lugbara	Nilo-Saharan	Central Sudanic	
Mohawk	Iroquian	Northern Iroquoian	1st and 2nd person unmarked for number
Paipai	Hokan	Esselen-Yuman	
Eastern Pomo	Hokan	Northern Hokan	
Turkish	Altaic	Turkic	
Tzutujil	Mayan	Quichean-Mamean	

b. Person: 1in/1ex/2/3
Number: sg/pl 23 languages

LANG	Family	Subfamily	Notes
Aceh	Austronesian	Malayo-Polynesian	Sg/pl in 1st exclusive only, no number distinction in 1st inclusive, 2nd, or 3rd
Marshallese	Austronesian	Malayo-Polynesian	
Chinese, Mandarin	Sino-Tibetan	Chinese	
Cubeo	Tucanoan	Central T.	
Dakota	Siouan	Siouan Proper	
Dong	Daic	Kam-Sui	
Kalihna	Carib	Northern Carib	French Guiana, Guyana, Surinam, Venezuela
Kwakiutl	Wakashan	Northern Wakashan	Sg/pl in 1st exclusive only, no number distinction in 1st inclusive, 2nd, or 3rd
West Makian	West Papuan	North Halmahera	
Central Marghi	Afro-Asiatic	Chadic	
Maxakalí	Macro-Gé	Maxakalí	Sg/pl in 1st exclusive only, no number distinction in 1st inclusive, 2nd, or 3rd
Miskito	Musumalpan	--	Objective and Genitive case-marked pronouns unmarked for number;

			Nominative pronouns distinguish singular and plural in all persons… except 1st inclusive, which has plural only
Central Sierra Miwok	Penutian	California	Singular inclusive form
Halh Mongolian	Altaic	Mongolian	No 3rd person pronouns, demonstratives used for this purpose
Ojibwa, Potowatomi	Algic	Algonquian	
Pakaásnovas	Chapacura-Wanha	Madeira	
Palauan	Austronesian	Malayo-Polynesian	
Quechua	Quechuan	Quechua 1	
Sirionó	Tupi	Tupi-Guarani	
Somali	Afro-Asiatic	Cushitic	
Telugu	Dravidian	South Central Dravidian	No 3rd person pronouns, demonstratives used for this purpose
Yaoré	Niger-Congo	Mande	
Yatzachi Zapoteco	Oto-Manguean	Zapotecan	

c. Person: 1/2/3
Number: sg/pl/du 10 languages

LANG	Family	Subfamily	Notes
Awtuw	Sepik-Ramu	Sepik	
Arapesh	Torricelli	Kombio-Arapesh	Dual in 1st person only
Hmong Njua	Hmong Mien	Hmongic	
Lithuanian			
Tonkawa	Coahuiltecan		
Tunica	Gulf		Dual in 3rd person masculine only
Wappo	Yuki		Dual in 3rd person only
Zuni	Isolate		No separate dual form for 1st & 2nd person nominative & accusative pronouns – dual number expressed via verb agreement; separate 1st and 2nd dual genitive pronoun, as well as 3rd nom., acc. and genitive
Navajo	Na-Dene	Nuclear Na-Dene	Obviative 4th person[47]
Central Yupik	Eskimo-Aleut		Obviative 4th person

d. Person: 1in/1ex/2/3
Number: sg/pl/du 7 languages

LANG	Family	Subfamily	Notes
Chinook	Penutian	Chinookan	
Comanche	Uto-Aztecan	Northern Uto-Aztecan	
Dieri	Australian	Pama-Nyungan	
Djingili	Australian	West Barkly	
Ngandi	Australian	Gunwingguan	
Ho	Austro-Asiatic	Munda	
Nama	Khoisan	Southern Africa	

e. Person: 1/2/3
Number: sg/pl/du/trial-paucal 1 language

LANG	Family	Subfamily	Notes
Yimas	Sepik-Ramu	Nor-Pondo	No 3rd person pronouns, demonstratives used for this purpose

f. Person: 1in/1ex/2/3
Number: sg/pl/du/trial-paucal 2 languages

LANG	Family	Subfamily	Notes
Fijian	Austronesian	Malayo-Polynesian	
Tok Pisin	Creole		

g. Person: 1/2/3
Number: (sg/pl/du) 1 language

LANG	Family	Subfamily	Notes
Kiowa	Kiowa-Tanoan	Kiowa-Towa	No 3rd person pronouns, demonstratives used for this purpose. Pronominal prefixes on verb encode person and sg/pl/du/ number.[48]

h. Person: 1/2/3
Number (sg/pl) 1 language

LANG	Family	Subfamily	Notes
Kutenai	Isolate		Word forms for 1st and 2nd person only, showing number distinctions[49]; clitics for all three persons unspecified for number.

i. Person: 1in/1ex/2/3
Number: none (sg/pl) 2 languages

LANG	Family	Subfamily	Notes
Perené Campa	Arawakan	Maipuran	m/f gender in 3rd person. Plural suffix -payeeri can attach to nouns and pronouns but it is infrequently used
Mixteco Chalcatongo	Oto-Manguean	Mixtecan	Pronominal prefixes on verb encode plural number.

References

Anderson, Stephen R. 1992. A-morphous morphology. Cambridge: Cambridge University Press.

Archangeli, Diana. 1988. Underspecification in phonology. Phonology 5.183-207.

Avery, Peter, and Keren Rice. 1989. Segment structure and coronal underspecification. Phonology 6.179-200.

Bejar, Susana. 2000. Structural markedness in formal features: Deriving interpretability. Revue québecoise de linguistique 28.47-72

Benveniste, Emile. 1971. The nature of pronouns. In: Problems in general linguistics. 217-222. Coral Gables, FL: University of Miami Press.

Berman, Ruth A. 1985. The acquisition of Hebrew. In Slobin, 255-372.

Blake, Barry J. 1994. Case. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1933. Language. New York: Holt, Rinehart and Winston. [Reprinted 1984. Chicago: University of Chicago Press.]

Boas, Franz. 1911a. The Kwakiutl Indian language. In Boas 1911b, 427-557. Washington. [Reprinted 1971. Seattle, WA: The Shorey Book Store.]

Boas, Franz. 1911b. Handbook of American Indian languages. Washington.

Brown, Cynthia A. 1997. Acquisition of segmental structure: Consequences for speech perception and second language acquisition. Montreal: McGill University dissertation.

Brown, Penelope, and Stephen C. Levinson. 1987. Politeness: Some universals in language usage. Cambridge: Cambridge University Press.

Brown, Roger. 1973. A first language: the early stages. Cambridge, MA: Harvard University Press.

Buehler, Karl. 1934. Sprachtheorie: Die Darstellungsfunktion der Sprache. Jena: Fischer

Campbell, Russell N. 1969. Noun substitutes in Modern Thai: A study in pronominality. The Hague, Paris: Mouton.

Camacho, José. 1996. Comitative coordination in Spanish. Aspects of Romance Linguistics: Selected papers from the Linguistic Symposium on Romance Languages XXIV, March 10-13 1994, ed. by Claudia Parodi, Carlos Quicoli, Mario Saltarelli, & Maria Luisa Zubizarreta, 475-488. Washington, DC: Georgetown U Press.

Carstairs-McCarthy, Andrew. 1987. Allomorphy in inflexion. London: Croom Helm.

Chiat, Shulamuth. 1978. The analysis of children's pronouns: an investigation into the prerequisites for linguistic knowledge. London: University of London dissertation.

Chiat, Shulamuth. 1986. Personal pronouns. Language acquisition. ed. by Paul Fletcher and Michael Garman, 2nd ed., 339-355. Cambridge: Cambridge University Press.

Clark, Eve V. 1985. The acquisition of Romance, with special reference to French. In Slobin, 687-782.

Clements, George N. 1985. The geometry of phonological features. Phonology Yearbook 2.225-252.

Cole, Peter. 1987. Null objects in universal grammar. Linguistic Inquiry 18.597-612.

Cooke, Joseph R. 1968. Pronominal reference in Thai, Burmese, and Vietnamese. Berkeley: University of California Press.

Corbett, Greville G. 1991. Gender. Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. Number. Cambridge: Cambridge University Press.

Cowan, H. K. J. 1955. Notes on Windesi grammar. Oceania 26.42-58.

Croft, William. 1990. Typology and universals. Cambridge: Cambridge University Press.

Cysouw, Michael Alexander. 2001. The paradigmatic structure of person marking. Nijmegen: Katholieke Universiteit dissertation.

Dalrymple, Mary, and Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. Language 76.759-798.

Déchaine, Rose-Marie. 1999. What Algonquian morphology is really like: Hockett revisited. Papers from the workshop on structure & constituency in Native American languages: MIT Occasional Papers in Linguistics 17, ed. by Leora Bar-el, Rose-Marie Déchaine & Charlotte Reinholtz, 25-72. Cambridge, MA: MIT Working Papers in Linguistics, Dept. of Linguistics, MIT.

Dixon, Robert M. W. 1977. A grammar of Yidiɲ. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 1988. A grammar of Boumaa Fijian. Chicago: The University of Chicago Press.

Durie, Mark. 1985. A grammar of Acehnese: On the basis of a dialect of North Aceh. Dordrecht: Foris Publications.

Everett, Daniel. 1986. Pirahã. Handbook of Amazonian languages, ed. by Desmond C. Derbyshire and Geoffrey K. Pullum, Volume 1.200-325. Berlin: Mouton de Gruyter.

Feuer, H. 1980. Morphological development in Mohawk. Papers and Reports on Child Language Development 18.25-42.

Foley, William A. 1986. The Papuan languages of New Guinea. Cambridge: Cambridge University Press.

Foley, William A. 1991. The Yimas language of New Guinea. Stanford: Stanford University Press.

Forchheimer, Paul. 1953. The category of person in language. Berlin: Walter de Gruyter.

Gazdar, Gerald and Geoffrey Pullum. 1982. Generalized Phrase Structure Grammar: A theoretical synopsis. Bloomington, IN: Indiana University Linguistics Club.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag. 1985. Generalized Phrase Structure Grammar. Oxford: Basil Blackwell.

Ginneken, Jacobus van. 1907. Principes de linguistique psychologique. Paris.

Grasserie, Raoul de la. 1888. De la véritable nature du pronom. Etudes de grammaire comparée. Louvain.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. Universals of language, ed. by Joseph H. Greenberg, 73-113. Cambridge, MA: MIT Press.

Gregores, Emma and Jorge A. Suárez. 1967. A description of colloquial Guarani. The Hague: Mouton & Co.

Guéron, Jacqueline. 1984. Inalienable possession, PRO-inclusion and lexical chains. Grammatical representation, ed. by Jacqueline Guéron, H.-G. Obenauer and Jean-Yves Pollock, 43-86. Dordrecht: Foris.

Halle, Morris. 1997. Distributed Morphology: Impoverishment and fission. PF: Papers at the interface: MIT Working Papers in Linguistics 30, ed. by Benjamin Bruening, Yoonjung Kang, and Martha McGinnis, 425-449. Cambridge, MA: MITWPL.

Halle, Morris and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. The view from building 20, ed. by Kenneth Hale and Samuel Jay Keyser, 111-176. Cambridge, MA: MIT Press.

Hagman Roy S. 1977. Nama Hottentot grammar. Bloomington: Indiana University.

Hanson, Rebecca. 2000. Pronoun acquisition and the morphological feature geometry. Calgary Working Papers in Linguistics 22, ed. by Susan Atkey, Jana Carson, and Michael Dobrovolsky, 1-14. Calgary: Dept. of Linguistics, University of Calgary.

Harley, Heidi. 1994. Hug a tree: Deriving the morphosyntactic feature hierarchy. Papers on phonology and morphology: MIT Working Papers in Linguistics 21, ed. by Andrew Carnie and Heidi Harley, 289-320. Cambridge, MA: MIT Working Papers in Linguistics, Department of Linguistics, MIT.

Harley, Heidi and Elizabeth Ritter. 2002. Structuring the bundle: a universal morphosyntactic feature geometry. Pronouns: Features and representation, ed. by Horst Simon and Heike Weise. 23-39. Amsterdam: John Benjamins.

Heath, Jeffrey. 1978. Ngandi grammar, texts, and dictionary. Canberra: Australian Institute of Aboriginal Studies.

Hockett, Charles F. 1966. What Algonquian is really like. International Journal of American Linguistics 32.59-73.

Hoff, Berend Jacob. 1968. The Carib language: Phonology, morphophonology, morphology, texts and word index. The Hague, Paris: Mouton.

Hoijer, Harry. 1933. Tonkawa: An Indian language of Texas. New York: Columbia University Press.

Huxley, Renira. 1970. The development of the correct use of subject personal pronouns in two children. Advances in psycholinguistics, ed. by Giovanni B. Flores d'Arcais and Willem J. M. Levelt, 141-165. Amsterdam: North-Holland.

Jakobson, Roman. 1971. Shifters, verbal categories, and the Russian verb. Selected writings, II: Word and language, 130-47. The Hague: Mouton.

Jespersen, Otto. 1924. The philosophy of grammar. London. [Reprinted 1992. Chicago: The University of Chicago Press.]

Joseph, Brian D. 1979. On the agreement of reflexive forms in English. Linguistics 17.519-523

Kimball, Geoffrey D. 1991. Koasati grammar. Lincoln, NE: University of Nebraska Press.

Kuroda, S.-Y. 1965. Generative grammatical studies of the Japanese language. Cambridge, MA: MIT dissertation.

Lehmann, Christian. 1988. On the function of agreement. Agreement in natural language, ed. by Michael Barlow and Charles A. Ferguson, 55-66. Stanford: CSLI.

Lipkind, William. 1945. Winnebago grammar. New York: King's Crown Press.

Loving, Richard. 1973. An outline of Awa grammatical structure. The languages of the eastern family of the East New Guinea Highland stock, ed. by Howard McKaughan, 65-87. Seattle: University of Washington Press.

McNally, Louise. 1993. Comitative coordination: A case study in group formation. Natural Language and Linguistic Theory 11.347-379

Mithun, Marianne. 1989. The acquisition of polysynthesis. Journal of Child Language 16.285-312.

Morgan, Lawrence Richard. 1991. A description of the Kutenai language. Berkeley, CA: University of California, Berkeley dissertation.

Murane, Elizabeth. 1974. Daga grammar: From morpheme to discourse. Norman, OK: Summer Institute of Linguistics

Newman, Stanley. 1944. The Yokuts language of California. New York: The Viking Fund.

Noguchi, Tohru. 1995. The role of syntactic categories in anaphora. Amherst, MA: University of Massachusetts dissertation.

Noyer, Robert Rolf. 1992. Features, positions, and affixes in autonomous morphological structure. Cambridge, MA: MIT dissertation.

Noyer, Rolf. 1997. Features, positions, and affixes in autonomous morphological structure. New York: Garland Press.

Petitto, Laura A. 1987. On the autonomy of language and gesture: Evidence from the acquisition of personal pronouns in American Sign Language. Cognition 27.1-52.

Pingkarawat, N. 1985. Distribution and reference of empty pronouns in Thai. Ms. University of Illinois, Urbana.

Plank, Franz and Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45. Linguistic Typology 1.53-101.

Popovich, Harold. 1986. The nominal reference system of Maxakalí. In Wiesemann, 351-358.

Ramstedt, G.J. 1939. A Korean grammar. Helsinki: Suomalais-Ugrilainen Seura.

Rice, Keren, and Peter Avery. 1995. Variability in a deterministic model of language acquisition: A theory of segmental elaboration. Phonological acquisition and phonological theory, ed. by John Archibald, 23-42. Hillsdale, NJ: Lawrence Erlbaum.

Ritter, Elizabeth. 1997. Agreement in the Arabic prefix conjugation: Evidence for a non-linear approach to person, number and gender features. Proceedings of the Canadian Linguistic Association Conference, ed. by Leslie Blair, Christine Burns and Lorna Rowsell, 191-202. Calgary: Dept. of Linguistics, University of Calgary.

Ritter, Elizabeth, and Heidi Harley. 1998. Meaning in morphology: A feature-geometric analysis of person and number. Paper presented at GLOW 20, Tilburg, The Netherlands.

Rood, David S. 1976. Wichita grammar. New York: Garland Publishing Inc.

Sagey, Elizabeth. 1986. The representation of features and relations in non-linear phonology. Cambridge, MA: MIT dissertation.

Schieffelin, B. B. 1985. The acquisition of Kaluli. In Slobin, 525-594.

Schmidt, Pater Wilhelm. 1919. Die Personalpronimina in den australischen Sprachen. Denkschriften der österreichischen Akademie der Wissenschaften. Vienna: Philosophisch-historische Klasse, vol. 64.

Schwartz, Linda J. and Timothy Dunnigan. 1986. Pronouns and pronominal categories in Southwestern Ojibwe. In Wiesemann, 285-307.

Showalter, Catherine. 1986. Pronouns in Lyele. In Wiesemann, 205-216.

Silverstein, Michael. 1985. Hierarchy of features and ergativity. Features and projections, ed. by Pieter Muysken and Henk van Riemsdijk, 163-233. Dordrecht: Foris.

Slobin, Dan Isaac (ed.). 1985. The crosslinguistic study of language acquisition. Volume 1: The data. Hillsdale, NJ: Lawrence Erlbaum.

Steinhauer, Hein. 1985. Number in Biak: Counterevidence to two alleged language universals. Bijdragen tot de Taal-, Land- en Volkenkunde 141: 462-485.

Sutcliffe, Edmund F. 1936. A grammar of the Maltese language. London: Oxford University Press.

The University of Calgary pronouns database (http://www.ling.ucalgary.ca/research/pronouns/index.html)

The University of Konstanz universals archive (http://ling.uni-konstanz.de/pages/proj/sprachbau.htm)

Thomason, Sarah G. and Daniel L. Everett. In press. Pronoun borrowing. Proceedings of the Berkeley Linguistics Society 27. Berkeley: Berkeley Linguistics Society, UC Berkeley

Wallace, Judith M. 1966. Katu personal pronouns. Mon-Khmer Studies II, ed. by. David D. Thomas et al, 55-62. (Publication No. 3 of the Linguistic Circle of Saigon, in collaboration with SIL).

Watkins, Laurel J. 1984. A grammar of Kiowa. Lincoln and London: University of Nebraska Press.

Westrum, Peter N. & Ursula Wiesemann. 1986. Berik pronouns. In Wiesemann, 37-45.

Willie, Mary Ann. 1991. Navajo pronouns and obviation. Tucson, AZ: University of Arizona doctoral dissertation.

Wiesemann, Ursula (ed.). 1985. Pronominal systems. Tübingen: Gunter Narr Verlag.

Wundt, Wilhelm. 1911. Die Sprache. Leipzig: Kröner-Engelmann

Zwicky, Arnold. 1977. Hierarchies of person. Papers from the thirteenth meeting, Chicago Linguistics Society, ed. by W. A. Beach, S.E. Fox and S. Philosoph, 714-733. Chicago: Chicago Linguistics Society.

*We would like to acknowledge our great debt to the fieldwork done by many linguists which enabled us to have access to the paradigms of a wide range of endangered languages. A number of the languages we cite have fewer than 20 speakers. For extensive comments on earlier versions of this work, thanks especially to Hotze Rullmann, Heike Weise and Horst Simon, Brian Joseph, and two anonymous reviewers. For other help with the development of these ideas, thanks to Diana Archangeli, Hagit Borer, Andrew Carnie, Eugene Chan, Michael Cysouw, Dan Everett, John Frampton, Diana Green, Ken Hale, Morris Halle, Riny Huybregts, Eloise Jelinek, Kyle Johnson, Mary Laughren, Alec Marantz, Rolf Noyer, Phoevos Panagiotidis, Frans Plank, Harold Popovich, Keren Rice, Leslie Saxon, Margaret Speas, and audiences at the University of Victoria, Ben Gurion University of the Negev, Tel Aviv University, the University of Toronto, the 1997 Annual Meeting of the Canadian Linguistics Association, GLOW 98, and the special session on pronouns at the DGfS 2000 meeting in Marburg. Thanks as well to the research assistants on the University of Calgary Pronoun Database project, Rebecca Hanson, Heather Bliss, Jen Abel, and Sheena van der Mark. The material in sections 1, 3 and 4 of this paper represents a significantly expanded and revised version of ideas presented in Harley and Ritter (2002).This research was supported by SSHRCC grant #410-98-0328 to Ritter and an NSF Postdoctoral Fellowship to Harley. The ordering of the author's names is alphabetical; this work represents equal effort on the part of both authors.

[1] Since agreement markers are realizations of the same morphological features, we assume that this set of morphemes will also lend itself to a geometric representation. However, there are a number of issues to be sorted out before our proposal can be extended to this domain. Notably, we need a better understanding of the nature of the grammatical mechanism involved in

agreement (copying vs. checking, for example), as well as a reliable diagnostic for distinguishing between pronominal clitics and verb agreement.

[2] Some systems assume that some features are mutually exclusive, while others are not; so, for example, Dalrymple and Kaplan (2000) make conjoined use of a [Speaker] and [Hearer] feature (as we do, see below) to capture inclusive forms, but assume that it would not be possible to simultaneously bear, for instance, [+fem] and [+masc]. This only adds to the general indeterminacy of the system.

[3] In approaches that do not make use of attribute-value matrices, the mutual exclusivity of proposed features such as [+sg] and [+dual] is sometimes explicitly treated by appealing to semantic incompatibility (Noyer 1992:55).

[4] Joseph (p.c.) points to the existence of first person comitative constructions like the Russian my s Borisom idjom, 'we with Boris are going' = 'Boris and I are going', where the plurality of the first person pronoun is at logical odds with its apparent referent, although not its containing DP. See the structural proposals in, e.g., Camacho (1996) and McNally (1993) for discussion.

[5] An inclusive singular consists of the speaker and exactly one addressee. For example, Ngandi, an Australian language, has singular, dual and plural numbers for all persons. In the case of the inclusive, singular, dual and plural specify the number of addressees included in the reference of the pronoun, rather than the total number of individuals (cf. Heath 1978). See also the discussion of Kalihna in section 2.4

[6]Forchheimer (1953:4-7) distinguishes two schools of thought in the literature. One position, which he notes is espoused by de le Grasserie (1888:3), van Ginneken (1907:211) and Boas

(1911b:39-40), groups 2nd and 3rd person in opposition to 1st, that is, assumes that the fundamental person dichotomy is between speaker and non-speaker. The alternative position, which Forchheimer attributes to Wundt (1911:141 ff.), Schmidt (1919:203), Jespersen (1924:212 f.), Buehler (1934:113) and Bloomfield (1933:252-3), is the one we adopt here, according to which the fundamental person dichotomy is that between participant in the speech act (1st and 2nd persons) and non-participants (so-called 3rd person). (This is also the position adopted by Jakobson 1971).

[7]Similar assumptions underlie Noyer's (1992) theory of person features. He reasons as follows (146):

> The interpretation of person features must recognize certain discourse roles as primitives: these roles are among the deictic markers of a speech-act, situating the speech-act with respect to its place, time and, in the case of person features, its participants. The primary distinction is between participants in the speech-act and non-participants, what Hockett (1966) refers to as "local" vs. "non-local" arguments. The speaker and [addressee] are local participants, while other parties, neither speaker nor [addressee], are nonlocal nonparticipants.

[8] Showalter (1986:206) notes that some mass nouns may form plurals, and that the pronoun used to refer to these plural groups is *be*.

[9] Many familiar Indo-European languages fit this description as well, with the additional complication of some gender distinctions, usually in the third person.

[10] Note that if Minimal is the default value for Individuation, then plural is always a marked category. This corresponds to Greenberg's (1963:94) Universal 35, which states that plurals are always marked in at least some cases. The same obtains for Speaker and Participant above, although languages that lack a plural number are relatively more common than languages without a 2nd person, presumably for external reasons. Note that the two are structurally comparable in that plural number requires activation of the Group node dependent on Individuation, and 2nd person requires activation of the Addressee node dependent on Participant.

[11] Leslie Saxon (p.c.) rightly points out that the intended interpretation of the combination of our "Minimal" and "Group" features is intersective, or restrictive modification, while the apparent interpretation of the combination of our "Speaker" and "Addressee" features is simple conjunction.

[12] The reverse pattern is simply ungrammatical (David Shaul and Ken Hill, p.c.).

[13] Chinook makes gender distinctions of masculine, feminine and neuter only in 3rd person singular; we have included only the feminine form here, since we are not dealing with gender at the moment.

[14] Corbett (2000:39) makes a distinction between determinate and indeterminate numbers. Singular, dual and trial denoting exactly one, two or three individuals, respectively, are determinate numbers; paucal and plural are indeterminate ones. We have not explicitly encoded this distinction in the geometry, with the advantage that it enables us to deal straightforwardly with languages that allegedly have a paucal without a dual, such as Bayso (Cushitic) or Walapai

(Yuman), cf. Corbett (2000:22) and references cited therein. In these languages, the paucal denotes between two and six individuals, rather than the usual case of three to six. We propose that a dual is a simply a determinate minimal group, and that the paucal in Bayso or Walapai is an indeterminate one, represented by the same Minimal Group geometry as the dual.

[15] Note that we predict that no language has both a trial and a paucal number; they are in complementary distribution, representing determinate and indeterminate interpretations of the same geometric configuration, as for the Bayso or Walapai dual/paucal in fn. 14 above. The languages that we surveyed have maximally four distinct numbers (singular, dual, trial/paucal and plural); see section 1.3 and Appendix for details. See also Foley (1986) and Croft (1990) for a similar claim. Corbett (2000:22-50) also discusses some Oceanic languages, including Sursurunga, Lihir and Mele-Fila which reportedly have a fifth number, either a greater paucal or a greater plural. The existence of such systems could be accommodated in our framework by the addition of a node, probably as a dependent of Group. We leave this issue open to future research.

[16] In footnote 5 above, we noted the existence of rare languages with a singular, or more accurately, minimal inclusive pronoun referring to the speaker and exactly one addressee. In Section 2.4, we presented the analysis of one such language, Kalihna. More commonly we find that there is simply a gap in the singular cell of the inclusive pronoun paradigm. For example, all nine languages in our database that have both an inclusive person and dual number have a dual inclusive pronoun. In eight of these languages, there is not also a singular/minimal inclusive, and the dual refers to exactly two individuals, one of whom happens to be the speaker and the other the addressee. Thus, the treatment of an inclusive pronoun referring to exactly two individuals varies across languages, and choice between treating it as a singular/minimal or a dual seems to

depend on the presence or absence of other forms in the paradigm, and on morphological composition. Given that the morphological shape of the Boumaa Fijian 1st inclusive form shares elements with the other duals, and that there is a gap in this paradigm, we treat it as dual, rather than singular/minimal.

[17] In compiling the database, an attempt was made to represent every 'major' language family and/or branch in the world, where 'major' meant a family or branch of over twenty languages. Data for the Chibchan, Panoan, East Papuan, and Geelvink Bay families, and the Tungus (Altaic), Formosan (Austronesian), Armenian (Indo-European), Kordofanian (Niger-Congo) and Northeast Caucasian (North Caucasian) branches of their respective families, has not yet been gathered due to unavailability of source material.

[18] We have not reported our findings on Japanese, Thai and Vietnamese in the tables in the Appendix, as it is unclear whether the sources are describing personal pronouns or noun substitutes; see sections 5.2 and 6.2 for discussion. The only other language which has been omitted is Wichita (Caddoan). According to Rood (1976:10), Wichita has no independent pronouns, and inflected participles of the verb 'be' are used for this purpose.

[19] For the purposes of the tables here and below, we have combined figures for languages with personal and demonstrative pronouns, using 3 for both. We also analyzed the languages that do not express number on pronouns but do express it elsewhere (e.g. on nouns or verbs) as having number distinctions, for reasons that we elaborate in section 4.

[20] While the figures are generally as we predict, there is one perhaps surprising result. Given the discussion of markedness as node-counting in section 1.3, it seems that the geometry predicts that systems with dual numbers (three persons, three numbers) should be roughly as frequent as

systems with inclusive persons (four persons, two numbers), because both systems require the same number of nodes in their most complex representation. However, dual number appears to be less frequent than inclusive person (twenty-one languages with dual number compared to thirty-four languages with inclusive person). This is consistent with an overall pattern which may reflect other influences: for example, person is more salient than number. While there are languages, albeit few, with no number distinctions, it is not obvious that there are languages with no person distinctions. See discussion in Sections 4.1, 4.2 and especially 6.2 below.

[21] But see discussion in footnote 14.

[22] Presumably, the reverse situation could not exist, for reasons having nothing to do with feature geometry: a pronoun system cannot exist without 1st or 2nd person paradigms. However, see discussion of Thai and Japanese in Section 6.2.

[23] Diana Archangeli (p.c.) points out that there is a natural analogue to this situation in the phonological literature. Consider the case of a phoneme that is specified for [+low] and [+front] features simultaneously. The [+low] feature pushes the tongue body back, meaning that it conflicts somewhat with the [+front] feature. Depending on the structure of the rest of the phonological inventory, different languages will resolve this conflict in different ways. Simpler systems will take advantage of the empty phonetic space in varying directions, resulting in different phonetic realizations of the same geometry.

[24] In earlier work (Harley and Ritter, to appear, Ritter and Harley 1998), we argued that Berik was a numberless language that used the Speaker node for 1st plural in the absence of an inclusive/exclusive distinction. However, Corbett (2000:65), citing Westrum and Weisemann (1986), claims that while Berik does not mark number on its pronouns, it does mark agreement

for number and gender on the verb. If this is indeed the case, our treatment of its pronominal system will need to include these contrasts; an approach like that we take to Koasati (section 6.1 below), where Individuation is present but not realized by the independent pronouns, appears to be indicated.

[25] // denotes a lateral click in this paradigm.

[26] Moreover, neither the direct nor the inverse agreement form is acceptable when the non-inclusive argument is 2nd person. We attribute this ungrammaticality to a constraint against the overlapping syntactic reference of the 2nd person (excl) and the inclusive form, along the lines of Guéron's (1984:44) Nondistinctness Constraint.

[27] In his discussion of universal 45, Greenberg (1963:96) does not specify any categorial restrictions. However, it is worth noting that gender distinctions in the 1st singular are extremely rare. In a survey of 110 languages, we found only 4 that have any gender distinctions in the 1st singular, a fact we return to below.

[28] Plank and Schellinger (1997:72) note that similar exceptions are found in Lithuanian where 1st and 2nd person dual pronouns are marked for gender, but singular and plural are not. On closer inspection, we see that the Lithuanian dual pronouns are also bi-morphemic, consisting of a pronominal stem morpheme specifying person, and the numeral two, which is inflected for gender.

(i) Lithuanian (Plank and Schellinger 1997:71).

	singular	dual		plural
		masc	fem	
1st	àsÃ	mù-du	mù-dvi	meS Õ
2nd	tù	jù-du	jù-dvi	juS Õ

Plank and Schellinger (1997) discuss other violations of Greenberg's Universal 45 which are more problematic for our analysis. We return to these in section 6.3.2 below.

[29] Given analyses of morphological splitting like those of Noyer (1997) and Halle (1997), who treat the Afro-Asiatic Imperfect conjungation with a combination of Fission and Impoverishment mechanisms, it is perhaps possible to view bi-morphemic realizations of agreement features as instances of tree-splitting, perhaps motivated as a way to repair feature geometries that violate a universal or language-specific constraint.

[30] It is not an easy task to distinguish between items that belong to the closed class of pronouns and those that belong to the open class of noun substitutes in these languages. For example, Cole (1987:603), citing Pingkarawat (1985), provides the following examples, which show that khaw 'he' has the binding properties of a pronoun, rather than a noun.

(i) Chart$_i$ bòòk waa khaw$_i$ hen Nuan
Chart speak say he see Nuan
'Chart$_i$ said that he saw Nuan$_i$.'

(ii) *Khaw$_i$ bòòk waa Chart$_i$ hen Nuan
he speak say Chart see Nuan
'He said that Chart saw Nuan.'

Tadmor (p.c.) observes that khaw is only used pronominally. Nevertheless, one might argue that this is a noun substitute because it also seems to contain a significant amount of extra-grammatical semantic content. According to Cooke (1968), khaw is used as a 1st person singular pronoun, in an intimate situation, if the speaker is a young woman or child or if the addressee has a slightly inferior status relative to the speaker. It is used as a 3rd person singular pronoun if the individual referred to has a status similar to that of the speaker in a relatively formal situation, or more generally in a neutral (neither formal nor intimate) situation. See also Thomason and

Everett (2001) for discussion of the difference in borrowing possibilities for pronouns in open vs. closed systems.

[31] Actually, in some such cases, the realization of agreement is dependent on both the formal features and the actual intended referent of the form, as Joseph (1979:520) illustrates for the reflexive form of the 'nursely' we. In examples such as ‘We seem a bit displeased with ourself/*ourselves, don't we?’ the self-morpheme agrees in number with the addressee, that is, the patient, although the pronominal morpheme's features match those of we: '

[32]Lehmann (1988:56-8) distinguishes between internal and external agreement. The former is expressed on specifiers and modifiers inside the noun phrase, and never includes person. The latter is expressed outside the noun phrase, on governing verbs, nouns and adpositions, and never includes case.

[33] Bejar (2000) proposes that not just case but the entire geometry is constructed by the syntax, which moves and merges features to construct well-formed structures defined by the geometry we provide here.

[34]It will likely turn out that the case node will need to distinguish between structural cases, which are assigned to subjects and direct objects, and local or oblique cases, which are assigned to other noun phrases, i.e. between nominative, accusative, absolutive and ergative cases on the one hand, and oblique cases on the other. Silverstein (1985) notes that languages with nominal-based ergative splits assign ergative/absolutive or nominative/accusative case on the basis of properties of the argument itself. More specifically, he argues that choice among these cases for agents and patients in such languages can be predicted with the hierarchy in (i):

(i) (1&2)non-Sg. >(1&2)Sg > 3 > Proper Names > Human > Animate > Inanimate

The University of Konstanz Universals Archives lists a number of counter-examples to Silverstein's hierarchy, but significantly, they all involve reranking of the different classes of DPs he originally identified. As for local cases, it has been argued that these often evolve from nouns, verbs or adverbs; see Blake (1994). The spatio-temporal specification of the source could provide the conceptual content necessary for inclusion of these cases in a feature geometry.

[35] In our database of 110 languages, 63 languages have some gender distinctions in 3rd person (and/or demonstratives). Of these, 12 also have some gender distinctions in both 1st and 2nd person, and 7 also have some gender distinctions in 2nd person only.

[36] These are the only languages in our database of 110 languages with this property.

[37] The opposite system, with number only in the 3rd person, is also attested, but is completely unproblematic for our model. In Mohawk 1st and 2nd person pronouns lack a number specification, but most 3rd person pronouns consist of a set of items which are specified for both number (singular or plural) and gender (masculine, feminine or zoic) (Mark Baker, p.c.). In our framework, the 1st and 2nd person pronouns would require activation of only the Participant node, and the 3rd person pronouns would require activation of only the Individuation node.

[38] Note that the Koasati dual marking here could be interpreted as additional support for our treatment of dual number as the combination of two features, Group and Minimal: the suppletive singular verb stem a:y(a) is supplemented by the addition of a marker which Kimball glosses as ‘non-singular’; we interpret this added marker as a realization of the Group node.

[39] An analogous problem appears to arise in the pronominal systems of Arapesh, a Torricelli language, and Yidiɲ, an Australian language, which have a 1st person dual pronoun, but no 3rd person dual, again counter to the geometric prediction schematized in 31. These are the only two languages with this system that we know of. Dixon (1977:166), discussing Yidiɲ, makes special mention of the fact that the dual 1st person pronoun ɲali is extremely rare, and that the 1st plural form is normally used to denote groups of two, one of whom is the speaker. Such a system is not impossible to acquire, given that the learner would have positive evidence for the dual 1st person form. However, s/he would then 'expect' to hear a third person dual, which would not be forthcoming in the input. While the pronominal inventory of a language like Arapesh or Yidiɲ may be learnable, we predict that systems of this type, which do not instantiate a regular geometric progression of complexity, will be rare cross-linguistically and unstable over time. The fact that the Yidiɲ 1st dual pronoun is very rarely used, together with the fact that closely related languages of the Pama-Nyungan subfamily do not have a 1st dual pronoun, is at least consistent with this prediction.

[40] Kuroda (1965) argues that Japanese pronouns are better analyzed as nouns rather than functional or closed class items. Noguchi (1995) qualifies this claim, in that he argues that this is true of personal pronouns, but the language has a distal demonstrative (sore) which functions pronominally in that it can be used as a bound pronoun with inanimate reference. See also discussion of Thai in Section 5.2.

[41] Cysouw (2001) provides a comprehensive survey of person and number marking in pronoun and agreement paradigms, but makes no distinction between agreement and pronouns in his

treatment. He does, however, acknowledge that they may exhibit different characteristics. See also fn. 44.

[42] Corbett (2000:51) also cites Kawi (Old Javanese) and Classical Chinese as languages that have been reported to lack number.

[43] The notion of plurality is expressed in other ways, notably by conjunction or by the addition of an associative/comitative PP. Everett (1986:281) tentatively lists, in addition to a 3rd person pronoun unspecified for number and gender, one which is specified as feminine (xi) and another which is nonhuman (xís). However, Thomason and Everett (in press) argue that these are not true pronouns, as they may appear only when attached to another word, and are used only contrastively, to emphatically distinguish the feminine or non-human status of the referent. The true free third-person pronoun hi has specifically masculine reference only in contrast with the xí affix.

[44] Michael Cysouw (p.c.) informs us that Kawesqar (Alcalufan) has a pronoun caw for 2nd and 3rd person. We predict that either the persons are disambiguated by verb agreement, or that this is an accidental homophony, occurring in one case of the paradigm, but not elsewhere. Cysouw (2001) notes that syncretism in agreement paradigms is quite common, and that such syncretisms may involve any combinations of persons. Notably, however, syncretisms of person are virtually unattested in paradigms of (overt) pronouns.

[45] The University of Konstanz Universals Archive, a comprehensive archive of statements of linguistic universals, lists no exceptions to this claim.

[46] Our intuition suggests that it might be the case that a geometry-disproving paradigm would also violate Carstairs-McCarthy's (1987) Paradigm Economy principle. Further investigation of the contrastive possibilities of feature-geometries is necessary to determine this.

[47] Mary Willie and Eloise Jelinek (p.c.) inform us that the so-called 4th person is better characterized as a type of 3rd person form, whose distribution is not thoroughly understood. It appears in some contexts to be behaving like an obviation marker, singling out referents who are not 'proximate' in the discourse (i); in others, it seems to have a politeness function, used in conjunction with kinship terms or in place of the first person when it is appropriate to 'distance' the speaker from a request or desire. In some cases, it seems to affect the aspectual interpretation of an embedded event, providing a type of imperfective reading.

i) Jáan Jein yił naalnish. Doo eii Séelii bił n**ji**lnishda.
John Jane 3.with 3.works Not Sally 3.with **4**.works
"John works with Jane. **He** doesn't work with Sally."

Whatever the correct treatment of these extremely interesting forms, we do not attempt to incorporate them into our geometry for the same reasons that we outline for our omission of formality considerations in section 5: the principles that govern the appearance of the 4th person are clearly context-dependent. For further discussion, see Willie (1991).

[48]The independent pronouns, which are uncommon optional elements in Kiowa, are not themselves specified for number; see Watkins (1984:100ff). However, this language has an elaborate system of number marking on nouns and verbs, which is used to identify a noun phrase as singular, plural or dual. In fact, the number system is quite complex, involving in addition to the expected number prefixes, an inverse marker for 3rd person. The latter expresses singular number for nouns that are basically dual or plural, and plural number for nouns that are basically

singular or dual. A pronominal system of this type could be handled in our framework, if we assume that the basic number is the underspecified option, lexically listed for each noun. The inverse number simply represents the marked alternative. For an insightful discussion of inverse number along these lines, see Corbett (2000:159-166).

[49] According to Morgan (1991:420), 'The independent pronouns in Kutenai are all syntactically third person nominal phrases, regardless of any apparent first or second person reference.' First and second person independent pronouns are comprised of a first/second person possessive proclitic and a third person nominal base. Note that the number distinction in 1st and 2nd person is therefore not part of the pronominal form, but marked on the 3rd person base.