*ABA and The Combinatorics of Morphological Features

Jonathan David Bobaljik[*]
University of Connecticut & Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS)
Storrs, CT, USA & Berlin, Germany

Uli Sauerland[†‡]
Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS)
Berlin, Germany

DRAFT, February 2017

Abstract In several three cell paradigms, it has been observed that one logically conceivable pattern – ABA under some arrangement of cells – is unattested. Existing approaches assume that such *ABA generaizations provide evidence for feature inventories which are restricted to features that stand in containment relations. We present a novel approach to *ABA generalizations that derives from general properties of feature-based morphology. To this end, we develop a formal account of the widespread view that morphological paradigms derive from ordered rules that relate abstract features from an inventory to morphological exponents. We demonstrate that without any further assumption the feature-based view restricts the space of typological patterns. We show furthermore that the feature-based theory derives *ABA generalizations not only from assuming a specific inventory and Pāṇinian rule order, but alternatively from the assumption that the inventory of features must be minimal if extrinsic rule order is allowed. Furthermore we discuss which explanation might be correct for actual cases of *ABA constraints, and we explore the consequences of the feature-based general approach for paradigms with more than three cells.

Keywords: features, morphology, combinatorics, syncretism, typology

* jonathan.bobaljik@uconn.edu
† uli@alum.mit.edu
‡ Authors are listed alphabetically – both authors are equally first authors.

1 Introduction

One of the most interesting and difficult questions in research on language lies in formally characterizing the class of possible grammars. One aspect of this challenge asks whether there are constraints on grammars of a general, abstract nature, and in turn, whether these constraints are specific to language or instantiations of even broader, domain-general constraints on cognitive systems, with manifestations observable elsewhere. For example, some progress has been made in syntax on the basis of *Formal Language Theory* and the Chomsky hierarchy (Chomsky 1956) for the analysis of sets of string sequences. We aim to contribute to the development of a similarly general perspective for morphology, particularly with respect to morphological features, i.e. the features that underlie the variation in how different concepts are grouped across languages as evidenced by exponence by the same form. The architecture of feature-based morphological systems predicts that only certain patterns of variation are possible. In this paper, we address *ABA generalizations from this perspective. We show that *ABA generalizations can be derived from the feature-based architecture in conjunction with a minimality assumption. We furthermore argue that such a derivation may be plausible for some cases of an *ABA generalization, but not for others.

The term **ABA generalization* refers to morphological patterns in which, given some arrangement of the relevant forms in a structured sequence, the first and third may share some property "A" only if the middle member shares that property as well. If the middle member is distinct from the first, then the third member of the sequence must also be distinct. Bobaljik (2012) demonstrates that a *ABA generalization holds for adjectival suppletion in the sequence positive-comparative-superlative: across a large cross-linguistic sample, one finds ABB patterns such as *good-better-best*, where the comparative and superlative share a root *be(t)-* distinct from the positive, but what is not found is an ABA pattern: **good-better-goodest*, in which the positive and superlative share a root, distinct from the comparative. Similar *ABA effects have been noted in extensive studies of case syncretism (Caha 2009), suppletion for both case and number in pronouns (Smith et al. 2016), Germanic verbs and participles (see Wiese 2008 on German, and class material cited by Starke 2009 on English), and in other domains.

In one way or another, existing accounts of these generalizations have argued that the *ABA effect arises as a result of nesting or containment relations among features, along with the assumption that linguistic rules are arranged such that a more specific rule takes precedence over (bleeds) a more general one, the so called Elsewhere or Pāṇinian ordering (Kiparsky 1973, 1979). For the example above, Bobaljik argues that the representation of the superlative properly contains the representation of the comparative, which in turn properly contains the basic form of the adjective, as in (1).

(1) a. Positive: [ADJECTIVE]
b. Comparative: [[ADJECTIVE] COMPARATIVE]
c. Superlative: [[[ADJECTIVE] COMPARATIVE] SUPERLATIVE]

If a language has a rule of suppletion such as GOOD $\mapsto$ *be(t)-* / __ COMPARATIVE, that rule will block the basic root *good* in both the comparative and the superlative, in virtue of being the most specific rule compatible with the context. Nothing forces the comparative and superlative to share a root – Latin uses an ABC pattern *(bonus-melior-optimus)* with a distinct root in each of the three grades, but the containment relation in (1) ensures that the ABA pattern is inderivable (except as a case of accidental homophony).

In this paper, we discuss some results of an ongoing project studying the combinatorial properties of rule systems that describe syncretism in morphological paradigms. Although that project did not set out to examine *ABA patterns per se, it turns out that *ABA emerges as a prediction in certain contexts, as a consequence of the assumption that Universal Grammar selects the minimal feature inventories needed to generate a paradigm of a given size. We believe this is interesting, since the *ABA restriction emerges without the containment/nesting hypothesis that characterizes other accounts. Intuitively, *ABA emerges when a three-element sequence is the product of two overlapping features and their intersection: in the sequence ("paradigm") $\langle x, y, z \rangle$, if x and y share a feature, and y and z share a feature, but x and z do not share a feature, then even without a total containment relation among the features, it follows that the patterns ABC, ABB, and AAB are generable, but ABA is excluded. Most of the paper is devoted to showing that this state of affairs is not only formally possible, but is in fact forced in some contexts by plausible minimal assumptions about feature logics. While this approach seems implausible for some *ABA patterns (we think there are good reasons independent of suppletion to assume that superlatives contain comparatives), we wish to bring this to the table as a possible alternative in other instances.

Although we have identified the *ABA result as an important point of contact with other current theoretical morphosyntax work, a significant portion of this paper will be devoted to presentation of a framework where classes of morphological models can be formally discussed, and where the effects of individual assumptions can be explicitly computed, for example, in terms of their restrictiveness. Alongside the *ABA result, we also discuss the effect of imposing Pāṇinian ordering on feature models, and show that its effects are comparatively weak in certain classes of model.

2 Paradigm-Partitions, Features, and Sequences

We start by recognizing that a *paradigm* is a list of cells, <*x,y,z,...*>, where each cell is a pairing of a linguistic form and a unique property or combination of properties. We take the order of the list to be arbitrary – paradigms in our view are descriptive devices and have no structure beyond that imposed by their constituent features.[1]

Syncretism is the observation that different cells in a paradigm may map to the same surface form.[2] Since we are interested in syncretism, we only care whether the cells of a paradigm are mapped to the same surface form or different ones. Abstracting away from the actual forms, x, y, z and so on, leaves us with a partition of a set of size n: the set of cells that may be mapped to linguistic forms for some linguistic entity, such as a given lexeme. The number of distinct partitions for an n-celled paradigm is the Bell number: B_n. For a three-celled paradigm, the $B_3 = 5$ distinct partitions are listed in (2):

(2) AAA, AAB, ABB, ABA, ABC

In medium- to large-scale studies of syncretism (Cysouw 2003; Bobaljik 2012; Baerman et al. 2005), it is commonly observed that only a subset, often only a very small subset, of the theoretically distinct partitions are attested. For example, Cysouw (2003, 2010) considers a sample of person paradigms in 250+ languages, characterized as an 8-cell paradigm space, but finds only 60-some-odd distinct partitions from among the logically possible $B_8 = 4,140$. The *ABA generalizations, mentioned above, make the same point: over some sizeable range of data, where 5 patterns are possible, only four are actually found

1 Technically, a list is ordered, and to present a paradigm on a printed page its cells must be ordered too. But this is a presentational necessity – the cells of a paradigm are not ordered, regardless of the one or two-dimensional arrangement we use to present it in the following. Technically, the general perspective views a paradigm as a mapping from a finite set $\{X,Y,Z,\dots\}$ of n elements to linguistics forms such as x, y, and z. In the following, we assume that there is a conventional enumeration of the elements of a paradigm $\{X,Y,Z,\dots\}$, and thereby the paradigm is equivalently the finite set is $\{1,2,\dots,n\}$. This assumption doesn't lose any generality and our approach is compatible with any one- or multi-dimensional arrangement of a paradigm.

2 An important, but in practice difficult, distinction to draw is the difference between accidental homophony and systematic syncretism; see Harbour (2008); Sauerland & Bobaljik (2013).

in the world's languages: AAA, AAB, ABB, and ABC, but not ABA. Typically, studies of syncretism seek explanations for such typological patterns — i.e. develop theories that predict only a subset of partitions to be possible. We address exactly this problem but one level of generality higher — we investigate how general assumptions about morphological analysis restrict which subsets of partitions can arise as typological predictions. For example, we show that a restriction to the partition-set {AAA, ABB, ABC} cannot be derived solely within our general assumptions, while the *ABA condition can be derived.

The general class of morphological models we explore is *feature-based* models. Such models have been prominent in morphological analysis, but their restrictiveness has not been investigated formally. At its most basic, a feature is a name for individual cells or sets of cells in a paradigm. With reference to an n-celled paradigm, we write a feature as f indexed with a binary vector, where 1 indicates the cell or cells that feature names. Thus, one way of naming features that generate a 3-celled paradigm is as in (3), with a unique feature naming each cell.

(3) a. f_{100}
b. f_{010}
c. f_{001}

We define a model of a given paradigm as having two components: an inventory of features, and Rules of Exponence, which relate features to form. Alongside the trivial feature inventory in (3), we may state the Rules of Exponence in (4) (here and throughout, capital letters are to be understood as variables ranging over phonological forms):

(4) a. $f_{100} \mapsto$ A
b. $f_{010} \mapsto$ B
c. $f_{001} \mapsto$ C

A model is grammar fragment, generating one paradigm. In the trivial example just considered, the model in (3) and (4) generates the paradigm <*A,B,C*>, a three-celled paradigm that is *maximally differentiated*, i.e., in which each cell has a distinct form.

Maximal differentiation is by no means the only way in which an n-celled paradigm space may be partitioned. Characterizing other partitions requires features that name (contain) more than one cell of the paradigm such as f_{110}. But when we specify rules of exponence for such features as well, more than one rule of exponence may be applicable to one cell, requiring the specification of rule order in the general case.

Two general logical operations on features are intersection and union (or conjunction and disjunction). Formally, both can be straightforwardly defined for the vector representation of features: the intersection of features f and f' contains a 1 in position m if and only if both f and f' have value 1 in position m. But the union of features f and f' contains a 1 in position m if and only if either f or f' or both contain a 1 in position m. We will assume below that conjunction (intersection) of features is available in language, while disjunction (union) is not.

For illustrative purposes, we show in the next paragraphs how the standard approach to the *ABA generalization in our current terms. As noted above, the model in (3)-(4) generates a single partition of the 3-celled space, namely <*A,B,C*>. In fact, from the inventory in (3), that is the only *complete* partition that may be generated. (By complete, we mean that a phonological form is assigned to every cell of the paradigm space.) Using only rules of exponence of the format in (4), only maximal differentiation is possible, because the cells share no features in common. Appeal to a "default" form implicitly invokes an additional feature, shared by all the cells: f_{111}, and there is no such feature in (4).[3]

The alternative inventory in (5) represents the standard approach to *ABA patterns in this notation:

(5) a. f_{001}
 b. f_{011}
 c. f_{111}

This encodes the same relationship among paradigm cells as in (1). One feature is shared by all three cells (this constitutes the default), one by two, and one is unique to a single element. On the assumption that the feature inventory in (5) remains constant across languages, but that the Rules of Exponence may vary from language to language or even from lexeme to lexeme, a variety of different paradigms (partitions) may be generated from this single, shared inventory of features. Rules of Exponence for two models sharing the inventory in (5) are given in (6) and (7).

(6) a. $f_{001} \mapsto \mathrm{C}$

3 Feature intersection is vacuous for the inventory in (3). If feature union were admitted, then any partition could be described from the inventory of features in (3) plus their unions. More generally, allowing feature union (as in Stump 2016) would have the effect that all partitions can be generated by any complete feature inventory.

b. $f_{011} \mapsto$ B
c. $f_{111} \mapsto$ A

(7) a. $f_{011} \mapsto$ B
b. $f_{111} \mapsto$ A

As the reader may verify, the model in (5) + (6) derives the maximally differentiated, ABC paradigm, one in which each cell is distinct from the others. The model consisting of (5) + (7) derives an ABB paradigm, with syncretism of the last two cells. In these models, rules of exponence are ordered sequentially (read by convention from top to bottom) – the first rule of exponence specified for any given cell must apply to that cell. In both (6) and (7), the final exponent (A) is the default – in principle it is compatible with all three cells – but it does not appear in those cells because the rule introducing the default is 'blocked' by the application of the more specific rules.

We represent the feature-based morphological analysis of a paradigm in a specific language more compactly as a *sequence* of features. (8-a) represents the ordered rules in (6) as a sequence and (8-b) that in (7). These examples also give the partition that each sequence generates.

(8) a. $\langle f_{001}, f_{011}, f_{111} \rangle$: ABC
b. $\langle f_{011}, f_{111} \rangle$: ABB

Some further examples of sequences are shown in (9). (9-a) and (9-b) are two sequences of the same features, but in the opposite order. (9-c) is a sequence of three features, but the last feature is redundant.[4] The sequence in (9-d) is incomplete – no rule of exponence assigns a feature to the third cell of the paradigm.

(9) a. $\langle f_{110}, f_{011} \rangle$: AAB
b. $\langle f_{011}, f_{110} \rangle$: ABB
c. $\langle f_{110}, f_{011}, f_{010} \rangle$: AAB *redundant*
d. $\langle f_{100}, f_{010} \rangle$: AB_ *incomplete*

In the following, we normally understand *sequence* to refer only to complete, redundancy-free sequences.

4 We can characterize redundancy abstractly as follows: In a feature sequence S the feature in position j is redundant, if the intersection of S_j with the union of the features $S_1, \dots, S_{j-1}$ is identical to S_j. If no feature in a sequence is redundant, we call it redundancy-free.

2.1 Pāṇinian Sequences

A point we return to in some detail below is that rule ordering may be extrinsic (a stipulated language-particular order) or intrinsic, i.e., such that more specific rules automatically bleed more general rules. A specific formulation of the intrinsic Pāṇinian ordering principle or Elsewhere Condition is as in (10) (after Kiparsky 1973:

(10) If two (incompatible) rules R1, R2 may apply to a given structure, and the context for application of R1 is a (proper) subset of the context for that of R2, then R1 applies and R2 does not.

This concept translates to our setup as follows:

(11) A (redundancy-free) sequence S is Pāṇinian if and only if any redundancy-free permutation of S yields the same partition as S.

Consider which of the sequences introduced in (9) satisfy (11). Since (9-a) and (9-b) are permutations of one another and yield different partitions, neither of them is Pāṇinian. (9-c) cannot be a Pāṇini-sequence since it isn't redundancy-free. But the redundancy-free sequence in (12-a) is Pāṇinian because it and its only redundancy-free permutation in (12-b) yield the same partition: ABC. The permutation (9-c) and three others contain a redundant feature.

(12) a. $\langle f_{001}, f_{011}, f_{110}\rangle$: ABC
b. $\langle f_{001}, f_{110}, f_{011}\rangle$: ABC

For Pāṇini-sequences the order of the sequence is redundant. Hence it is sufficient to represent Pāṇinian sequences as an unordered set, and we'll sometimes also use the term Pāṇinian set for a set S of features such that any redundancy-free sequence of all elements of S is Pāṇinian.

As example (9-a), the set of f_{110} and f_{011} illustrates, there are sets of features that don't allow any Pāṇinian sequences, but allow ordered-sensitive feature sequences. Now, all traditional analyses allow for Rules of Exponence to make use of intersection. Thus, if f_a and f_b are in an inventory, then $f_a \cap f_b \mapsto$ A is a well-formed Rule of Exponence. We adopt this wide-spread view in following, and assume that the intersection of features is generally available, as noted above. One consequence of this that, if a feature set allows the formation of a valid oder-sensitive sequence, than there is also a valid Pāṇinian sequence from that inventory. Specifically from f_{110} and f_{011}, a valid Pāṇinian sequence is $\langle f_{110} \cap f_{011}, f_{110}, f_{011}\rangle$. But note that this sequence doesn't generate the partition AAB as (9-a) does, but only the sequence ABC. For this reason (among others) we will examine in some detail below the consequences of assuming either that extrinsic ordering is allowed, or that grammars (models) are limited to Pāṇini sets.

As noted, we assume in the following that apart from intersection, no other algebraic operation on features is available.[5]

2.2 Partition Sets

We may now identify another concept that will be useful throughout this article. For any feature inventory $\boldsymbol{I}$, the *Partition Set* of $\boldsymbol{I}$ is the set of all partitions that may be generated from $\boldsymbol{I}$.

Generation of Partitions can either be understood as using extrinsically ordered sequences or as using only Pāṇinian sequences. We introduce the abbreviation OPS and PPS for these two concepts of Partition Set.

(13) a. The *Order-Partition-Set* (OPS) of Inventory $\boldsymbol{I}$ is the set of all partitions $\boldsymbol{P}$ such that there is an $\boldsymbol{m} \geq 1$ and a (complete, redundancy-free) sequence of features $\langle f^1, \ldots, f^m \rangle$ drawn from the closure of $\boldsymbol{I}$ under intersection that generates the partition $\boldsymbol{P}$.

b. The *Pāṇini Partition-Set* (PPS) of Inventory $\boldsymbol{I}$ is the set of all partitions $\boldsymbol{P}$ such that there is an $\boldsymbol{m} \geq 1$ and a Pāṇinian sequence of features $\langle f^1, \ldots, f^m \rangle$ drawn from the closure of $\boldsymbol{I}$ under intersection that generates the partition $\boldsymbol{P}$.

As we have shown, the OPS of (3) is just <ABC> and since inventory (3) is itself Pāṇinian, this is also its PPS. The discussion of (6) and (7) shows that the OPS of (5) contains (at least) ABC and ABB. Moreover, ABA is not in the partition set of (5). That is, regardless of rule order, no sequence using only the features in (5) will generate an ABA pattern from the inventory in (5) (except as an instance of accidental homophony). For inventory (5), the PPS is again the same as its OPS. In one way or another, the works cited above therefore take the absence of ABA patterns in the domains they investigate to indicate that the underlying features must be organized into the kind of monotonic containment relations represented in (5).

For the inventory of features in (12) the OPS and PPS differ. The only Pāṇinian sequence possible from these three features is the one shown in (12) generating the partition ABC. Hence the PPS contains only ABC. But the

5 The restriction to intersection is implicit in a great deal of formal morphological theory. There are approaches, such as Stump (2016), which allow other Boolean operators in the construction of complex features.

two non-Pāṇinian sequences $\langle f_{011}, f_{110}\rangle$ and $\langle f_{110}, f_{011}\rangle$ are both complete, and therefore the OPS of the inventory (12) is the set { ABC, AAB, ABB }.[6]

In this way, we see rather explicitly the general logic that relates typological generalizations to conclusions about features in universal grammar. In large scale investigations of syncretism in a given domain (case, or number, or adjectival gradation) the data we have are the attested partition sets. The *ABA generalization is a gap: in some domain, one logically possible partition is not contained in the attested partition sets. Feature-based explanations ask at one level what the feature inventory might be, such that ABA is not included in its paradigm set. At a second, higher level, we also ask whether feature-based explanations can predict any paradigm set as an OPS or PPS for some set of features.

In the following, we use the notation introduced here to explore the consequences of various kinds of formal restrictions one could conceivably apply to models of this sort (inventories of features and associated rules of exponence). We do so in the first instance entirely in the abstract, with no connection to substantive features or empirical data. Our goal is to better understand some of the formal properties of feature logics, and to compare the ways in which various intuitively plausible assumptions do and do not restrict the combinatorics.

2.3 Restricting the hypothesis space

One thing our notation calls attention to is that, absent any prior assumptions about the content of features, the number of possible features that can be defined grows quickly. For a paradigm of n cells, there are $2^n - 1$ non-empty features that may be defined. For a three cell paradigm, the 7 definable features are these:

(14) f_{100}
f_{010}
f_{001}
f_{110}
f_{101}
f_{011}
f_{111}

If features could be freely chosen to form inventories, then 128 distinct feature inventories could in principle be constructed from these features (including

6 It is generally the case that the PPS of an inventory must be a subset of its OPS.

the empty set). As we have seen above, from each inventory, a number of distinct models can be constructed. That is, each inventory can be mapped to one or more sequences, thus yielding a variety of partition sets. If rule ordering is unconstrained, then from a single inventory with n features, there are $n!$ distinct sequences that may be so constructed (although some number of these will be redundant). The number of possible models (and thus grammars) thus quickly becomes astronomical, and we suggest it is therefore important to ask whether there may be some universal constraints that drastically restrict the classes of possible models to be considered. Thus, we will spend a fair part of the discussion that follows discussing the combinatorics involved.

We will approach this as follows: Using the understanding of features, paradigms, and models outlined above, we will set out to explore in quantitative terms various conditions that may be imposed, and show explicitly how they do and do not restrict the space of possibilities. Many of the numerical results are non-obvious, and we provide the code in on-line supplemental materials for this paper.

Here, we define briefly the two conditions that will be central in the investigation that follows. Before proceeding to details, we show how various other conditions can be defined. Our formalism allows us, at least in principle, to selectively add or subtract conditions in order to be able to accurately examine the consequences of any particular set of assumptions.

2.4 Minimal Valid Inventory

We now define two formal conditions, which we suggest are a priori plausible conditions to restrict the class of possible inventories, and whose consequences we work through in detail below.

The first, basic condition is that an inventory be valid.

(15) An inventory I is *valid* for a paradigm P iff there exists a model M including I that generates the maximally differentiated partition of P.

The maximally distinct partition of a paradigm (space) is the partition in which each cell is distinct from every other cell. In other words, for a three cell paradigm, a valid inventory is one for which there is some set of rules that will derive ABC. Note that Validity is a property of inventories, not models (grammars). The Rules of Exponence in (6) demonstrate that (5) is a valid inventory, but we do not require that every grammar (model) generate ABC; syncretism by definition precludes there being such a requirement for every model.[7] The model constituting (5)+(7) is perfectly well formed (and well attested).

Validity is related to another condition: completeness, which we have mentioned above. Completeness is a condition on sequences (and thus derivatively on models):

(16) A sequence $\boldsymbol{S}$ is *complete* with respect to a paradigm $\boldsymbol{P}$ iff $\boldsymbol{S}$ generates a form (possibly zero) for every cell in $\boldsymbol{P}$.

The more interesting condition on inventories is Minimality:

(17) An inventory $\boldsymbol{I}$ constitutes a Minimal Valid Feature Inventory for some paradigm $\boldsymbol{P}$ iff

a. $\boldsymbol{I}$ is valid for $\boldsymbol{P}$, and
b. there is no alternative $\boldsymbol{I}'$ s.t., $\boldsymbol{I}'$ is also a valid inventory for $\boldsymbol{P}$ and $\boldsymbol{I}'$ has fewer features than $\boldsymbol{I}$

In the next sections, we will work through these conditions for various sizes of paradigms, starting with 2-celled and then 3-celled paradigms. One finding in this paper is that the two simple assumptions just noted – that inventories use the minimal number of features to describe a paradigm space – have the curious effect that in certain paradigm spaces, notably those with three cells, certain patterns of syncretism become unstatable. In a sense to be made clear below, ABA patterns of a certain type are indescribable. This result is of interest, because it arises without the nesting/containment assumption that plays a central role in other treatments of *ABA generalizations (Bobaljik 2012; Caha 2009; Starke 2009). Another result is a curious pattern in the nature of the restrictiveness that these assumptions create.

Because of the way we have defined features and inventories, intersection intersects with Minimality in a non-trivial fashion. Intersection, like rule ordering, allows for exponents that do not directly conform to features that are in the inventory. If an inventory consists only of f_{110} and f_{101}, a rule can be stated referring to: $f_{110} \cap f_{101} = f_{100}$. While this generates an exponent that only expresses the first cell, it does so without the feature f_{100} being contained in the inventory. There is nothing arcane about this: in effect it is similar to recognizing that a language can have an exponent for '[feminine,plural]' if it

7 Maximal differentiation is also not a requirement for every language. Famously, although there are many ways to define the case paradigms for Russian nominals, there is no paradigm that is maximally differentiated in Russian—all Russian case paradigms have some measure of syncretism (Jakobson 1936/1971; see Bobaljik (2002) for some implications of this old observation).

has the features [feminine] and [plural], without positing a special "feature" [fempl].

Before presenting the results, a few further remarks are perhaps in order about our notation and assumptions. The following subsection is an aside in terms of the logic of the argument to be given, but readers may find it useful in understanding the feature system a little more clearly. Any of a number of other conditions could be expressed in our system. The next paragraphs present some conditions on inventories that we could, but do not, explore in this paper. These remarks are to make explicit what we do not assume, alongside what we do, but also serve to show how our notation can expresses various common ideas in the literature.

2.5 Further conditions

Privativity v. Binarity

Our features are, by definition, privative, rather than binary, in the sense that these terms are understood in the morphological and phonological literature. Binary features, of the sort typically written [±F] are, in our terms, names for pairs of features: a feature that names a set of cells, and another feature that names the complement set. In our terms, feature binarity could be expressed by holding that if f_{1100} is a feature in some inventory, then f_{0011} is also a feature in that inventory, etc. We accord no special status to pairs of features in this way: an inventory containing f_{0011} may or may not contain the complement as a second feature; we impose no general restriction that it do so. (In at least some cases, including two- and three-cell paradigms, imposing binarity complicates the analysis).

Defaults

A default feature (or value) is one that is compatible in principle with all cells, i.e., $f_{111...}$. Like binary feature pairs, we consider full sets of inventories, including those that do and do not contain the default. The default feature has no special status at the outset – it is simply one feature among many to be considered. Again, we will largely allow the computation to determine whether accounts with default features are in some way more or less restrictive than corresponding accounts without.

Containment

Much of the ABA literature relies on partial or total containment among classes of features in an inventory. The Nanosyntax framework codes this as an *fseq*, assumed to be universal and invariant across languages (Caha 2009). Other ABA literature (Bobaljik 2012) assumes containment in the contexts where ABA is excluded, but without a total commitment to invariant *fseqs*.

In any event, we have seen how feature containment relations are expressed in our notation, as in (5). The *fseq* assumption would then elevate that to a general condition: for any two features f_a, f_b in an inventory, either $f_a \subset f_b$ or $f_b \subset f_a$. Once again, we do not impose a priori conditions of this sort, as our aim is to see whether these arise from other considerations.

3 Two-Cell Paradigms

Consider first the case of paradigms with two cells. Analysis of a two-cell paradigm space is relatively trivial, but serves as a warm up for the more interesting cases, and offers an opportunity to become more familiar with the notation for presenting the analysis.

For the analysis of a two-celled paradigm space, there are three logically possible features: f_{10}, f_{01}, and f_{11} – this corresponds to the general formula that for n-cells there are $2^n - 1$ possible features. From three features, eight distinct inventories of features may be defined, i.e., the power set of the features. Of these, we may discard the empty set - if there are no features, nothing can be described.

Of the seven remaining inventories, any inventory consisting of just a single feature will fail our criterion of Validity: The maximally differentiated partition of a two-celled paradigm space is AB, i.e., the two cells are distinct. Since our features are privative, a single feature is not sufficient to analyze the AB paradigm: If the single feature is f_{11} it isn't possible to make the required distinction between the A and the B cell – the only m-partition that could be analyzed is AA. And if the single feature was either f_{10} or f_{01}, no analysis of the two cell paradigm is possible at all. Only one cell could receive an exponent. Recall that we made the decision not to assign the 'default' f_{11} some special status but to include it as just one possible feature among many. Therefore if only a rule of exponence $f_{10} \mapsto A$ is specified, the second cell wouldn't be filled at all. Therefore this analysis fails to be valid under (15). This shows that at least 2 features are required to analyze the AB paradigm.

That two features are sufficient is shown by looking at Table 1. This table displays the four inventories with two or three features. For each inventory, the set of possible rules of exponence are given (redundant features are in parentheses), and in the rightmost column, the corresponding partitions that can be generated. As the table shows, any selection of two features from the three possible features will allow an analysis of the AB-paradigm. Thus these three subsets represent possibilities for a restrictive Universal Grammar

#	inventory	sequence	partition
1	f_{01}, f_{11}	f_{11}	AA
		$f_{11}, (f_{01})$	AA
		f_{01}, f_{11}	AB
		f_{01}	**
2	f_{10}, f_{11}	f_{11}	AA
		$f_{11}, (f_{10})$	AA
		f_{10}, f_{11}	AB
		f_{10}	**
3	f_{10}, f_{01}	f_{10}, f_{01}	AB
		f_{01}, f_{10}	AB
		f_{01}	**
		f_{10}	**
4	f_{10}, f_{01}, f_{11}	$f_{10}, f_{01}, (f_{11})$	AB
		$f_{01}, f_{10}, (f_{11})$	AB
		$f_{10}, f_{11}, (f_{01})$	AB
		$f_{11}, (f_{01}, f_{10})$	AA
		f_{01}	**
		f_{10}	**

Table 1 Table of valid feature inventories and corresponding partition sets for two-cells. ** = incomplete sequence

satisfying Minimality — the three-feature inventory (#4) is excluded by this criterion.

Recall from above that we defined a model as an inventory (set) of active features plus a sequence of rules of exponence. Thus each inventory in the table predicts multiple grammars (models) to be possible, each corresponding to some sequence of rules of exponence. The Table compresses each sequence (including incomplete ones or ones containing redundant features) to a single line.

Inventory #1 is valid, since there is a model containing this inventory, which generates the maximally differentiated partition AB. This model is in the third line: there are two, ordered rules of exponence ($f_{01} \mapsto$ B, and $f_{11} \mapsto$ A). As the table shows, the AA partition may also be generated from the same inventory. The first sequence provides a rule of exponence only for the feature

f_{11}. This generates the fully syncretic paradigm: AA. Continuing through the table, we see in this way that the first and the second possible universal inventory each allow two classes of languages corresponding to the partitions AA and AB. The third feature inventory, although Valid and Minimal, only predicts the AB partition as possibility. On this analysis AA would never surface, except as the result of accidental homophony.

The fourth inventory contains all three features and therefore allows 6 sequences with rules of exponence for 3 features, 6 sequences with 2 features, and 3 with single features, which we show in a condensed form in 1. However, as noted, this inventory fails the Minimality condition.

Typological evidence or learning experiments ultimately can inform us which partitions are attested. If a typological survey shows that both AA and AB patterns exist, then the the third inventory, though valid and minimal, is not the actual inventory made available by UG. We note in passing that it is the only minimal valid analysis that uses a binary feature, rather than a default and "marked" combination.

However, the typological evidence cannot alone decide between different inventories that both predict the same possible partitions like inventories #1 and #2 above.

Despite the relatively trivial nature of the exercise with the two-cell paradigm space, the preceding discussion demonstrates that assumptions have consequences, and the the assumption that UG inventories be both minimal and valid has reduced the space of possible inventories from 7 (or 8 with the empty set) to 3. We have shown how typological evidence can be brought to bear on the choice. Finally, we note that the two minimal valid inventories that are capable of generating both AA and AB patterns are in fact *linear permutations* of one another. Since we have taken the linear order of the paradigm cells to be arbitrary, there is no way to resolve this choice further on our assumptions.

Our first result is that in a paradigm space that constitutes only a binary opposition, the minimal valid analysis is one that takes UG to have a single feature that names one member of the opposition, and which is contrasted with a default feature, compatible with both members. In this way, there would be an empirically-grounded argument to be made that if Minimality is assumed, then Binarity should be rejected as a general condition on feature inventories. in the manner just noted, the two assumptions make contrasting predictions about the state of the world. But we have no way on these considerations alone of saying which member of the opposition is 'marked.'

Before moving on, we introduce one other aspect of notation that we will use again later in the paper. In the following we use a more compact and intuitive way to display inventories, possible valid sequences and the predicted

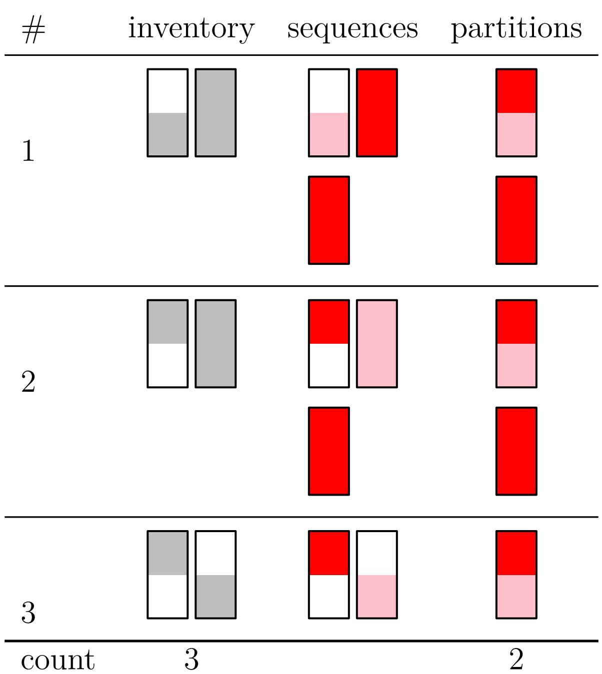


Table 2 Graphical display of table of minimal valid two-cell inventories and derivation of the predicted partitions.

partitions by representing the features, exponents, and partitions as vertical sequences of squares. In Table 2, we display the three valid systems that contain the minimal number of features, i.e., two features, in the two-cell case in this way. As we have already discussed, while there are three such distinct feature inventories, inventory 1 and inventory 2 predict the same sets of possible partitions. But inventory 3 predicts a smaller set of possible partitions, namely only the AB partition.

4 Three-Cell Paradigms

Turning to the three-cell paradigm space, we begin to see the growth in the space of analytical possibilities, and we also see how various assumptions such as Minimality and intrinsic, i.e., Pāṇinian ordering restrict that space. For a three-cell paradigm space, there are $2^3 - 1 = 7$ possible features, listed in (18):

(18) f_{100}
f_{010}

f_{001}
f_{110}
f_{101}
f_{011}
f_{111}

If features could be freely chosen to form inventories, then 128 distinct feature inventories could in principle be constructed from these features (including the empty set).

Validity

The first restriction we impose is Validity, as in (15). For example, the inventory $f_{100}, f_{010}, f_{001}$ is valid, in that it describes a three-way contrast, while the inventory $f_{100}, f_{110}, f_{010}$ is invalid - it is not complete, and thus provides no means to describe the third cell. It turns out that 96 of the 128 possible inventories of active features are valid in this sense in the three cell case (see table 3 below). With four cells, the ratio is 31 962 out of 32 768 (see table 5 below). Validity thus restricts the number of feature sets, but the restriction is not particularly strong.

Minimal Feature Inventory

The more interesting (and less obviously empirically motivated) requirement is Minimality. As defined above, a minimally valid feature inventory is an inventory that contains the minimal number of features needed to describe the maximally differentiated partition. For the two-cell space, the minimality requirement does not restrict the possibilities in any interesting way (it excludes only one inventory out of 8), but for the three-cell space, the minimal number of features that is needed to describe the maximally differentiated partition is two, as we show presently. Validity plus Minimality together thus restrict the choice from among 128 different logically possible feature inventories to the following three:

(19) a. f_{110}, f_{101}
b. f_{101}, f_{011}
c. f_{110}, f_{011}

No other combinations of two (or fewer) features generates the ABC array. (20) gives the rules of exponence that generate <*ABC*> from the inventory in (19-c), showing that two features are sufficient to satisfy Validity:

(20) a. $f_{110} \cap f_{011} = f_{010} \mapsto$ B
b. $f_{110} \mapsto$ A
c. $f_{011} \mapsto$ C

In order to describe the ABC pattern, the Rules of Exponence must be (partially) ordered, such that the exponent of the conjoined features takes preference over the rules in (20)b-c (this holds for any of the three inventories in (19)). The property of Order was not relevant in the two-cell paradigm, and so we make a brief digression to discuss it here.

Order

Because each of the inventories in (19) has two basic features that may be conjoined to define a third feature, the number of possible sequences for each inventory is 16, although many of these sequences will be redundant.[8] (20) illustrates one order. Other orders of the rules will yield syncretic partitions (paradigms). If the order of the first two rules in (20) is exchanged, then the rule in (21-a). will assign the exponent A to the first two cells, *bleeding* rule (21-b) (i.e., rendering rule (21-b) redundant), and yielding the partition AAC.

(21) a. $f_{110} \mapsto$ A
b. $f_{110} \cap f_{011} = f_{010} \mapsto$ B
c. $f_{0111} \mapsto$ C

As a general property (well understood from studies of Rule Ordering), ordering f_a before $f_a \cap f_b$ will render the conjunction redundant, and is thus equivalent to not selecting (or having no rule referencing) the conjoined feature. This corresponds to an *intrinsic* order: if the conjoined rule is active, it must be ordered before its individual conjuncts.

If the conjoined rule is omitted or not ordered first, then the order between the two rules referring to the basic features matters. Such ordering is *extrinsic* and must be stated explicitly. (22) yields AAC while (23) yields ACC. In both examples, the conjoined rule is ranked non-initially and these models are thus indistinct from corresponding models lacking the intersection rule.

(22) a. $f_{110} \mapsto$ A
b. $f_{011} \mapsto$ C
c. $f_{110} \cap f_{011} = f_{010} \mapsto$ B

(23) a. $f_{011} \mapsto$ C
b. $f_{110} \mapsto$ A
c. $f_{110} \cap f_{011} = f_{010} \mapsto$ B

The following table shows, for one inventory, the six possible models (six distinct orders of three rules) and the three corresponding partitions that are derived. (As before, redundant elements in the sequences are in parentheses).

8 The total number of arrangements of a set with n elements: $a(n) = n * a(n-1) + 1$, $a(0) = 1$.

The analogous table for the other two choices can be readily constructed. As an expository device, we use green text to indicate a feature that is derived as the intersection of the two basic features. As explained in Section 2.5, the green features are not part of the feature inventory, but are a convenient abbreviation for rules of exponence that make reference to the intersection of two features in their structural description.

(24)

inventory	sequences	partition
f_{110}, f_{011}	$f_{010}, f_{110}, f_{011}$	ABC
	$f_{010}, f_{011}, f_{110}$	ABC
	$f_{110}, (f_{010}), f_{011}$	AAC
	$f_{110}, f_{011}, (f_{010})$	AAC
	$f_{011}, (f_{010}), f_{110}$	ACC
	$f_{011}, f_{110}, (f_{010})$	ACC
		3

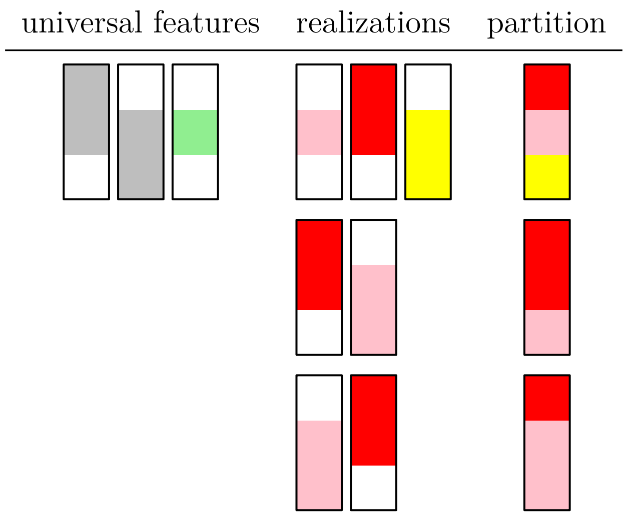
What (24) shows is the following: There are (only) three minimal valid feature inventories that can generate a maximally differentiated three-celled paradigm space. One such inventory is $\{f_{110}, f_{011}\}$. From that inventory, 6 (= 3!) sequences may be formulated, where each sequence is a distinct, total ordering of rules of exponence for the two features and their intersection.[9] While there are six rule orderings possible, only three distinct partitions are generated. The first two lines in (24) derive the same surface patterns (partitions), since the ordering of the last two rules is irrelevant.

As the reader may verify, the other two minimal valid inventories (in (19-a) have the same properties as (24). The three inventories amount to permutations in the order of the cells, but are otherwise identical in their formal properties.

The information in (24) is represented graphically in (25):

9 If non-total sequences are included, there are 15 possibilities, but the additional sequences are either incomplete, or indistinct from the sequences in (24) which have redundant rules.

(25)



4.1 Result: *ABA

At this point, we note two properties we believe to be of theoretical interest. For a three-celled paradigm space, there are $B_3 = 5$ distinct partitions. However, imposing the conditions of Validity and Minimality on the UG feature inventories restricts the expressive power of the system, such that each inventory generates only 3 of the 5 possible partitions, and the three so generated are moreover linear permutations of one another. We believe this is of interest since it appears to be true at least in some domains that the number of attested partitions is a small subset of the logically possible ones. The example we noted above was that in the 8 cell division of the person/number space, only 60-some-odd distinct partitions, out of $B_8 = 4,140$ possibilities, are attested in Cysouw's 250+ language sample. Being able to predict restrictions on the space of possibilities is thus of potential theoretical interest, if the restrictions indeed line up with the data. In the case at hand, the following restrictions obtain:

Of the five possible partitions of a three-cell spaced, four show some differentiation among the cells. However, each of the inventories in (19-a) generates only three of those partitions. As in the case of inventory #3 in the two-celled paradigms, we are now able to connect our formal results to potential empirical evidence. If there is, as we have hypothesized, a fact of the matter for some domain, such that UG contains only one of the inventories in (19-a), then this should show up as the following empirical generalization: across the relevant domain, only three of the four possible patterns of differentiated partition should be attested. In (24), we show that the inventory f_{110}, f_{011} generates the partition set <ABC, AAB, ABB>; that inventory does not generate ABA. No sequence from that inventory will generate a pattern in which the first and last cell share an exponent, to the exclusion of the middle cell. The same holds for the other two inventories, up to linear order: each inventory will fail to generate exactly one of the possible partly syncretic partitions.

This result is noteworthy in the current context, since it provides a means of characterizing the absence of *ABA patterns without assuming featural containment. Existing accounts of *ABA patterns (refs) are all built on what, in our terms, is a non-minimal feature structure, with strict nesting of features – some version of: $f_{100}, f_{110}, f_{111}$.

In other words, what we have just shown has two parts. The easy part is a demonstration that it is possible to derive a *ABA generalization for some domain without invoking containment. We have just done so. The slightly harder part was the demonstration that the type of feature inventory that derives *ABA without containment is not only possible, but is in fact preferred (over containment), if UG makes use of Minimally Valid feature inventories. We postpone until the next section some speculative remarks on whether this result constitutes a plausible alternative scenario for the account of *ABA generalization examples in the literature.

Before that discussion, we note one further point about these inventories. No valid, minimal inventory for a 3-cell paradigm space generates the maximally undifferentiated partition AAA. Curiously, it is not a general property of our assumptions that such undifferentiated partitions are universally excluded in the minimally valid inventories, rather, they are excluded for paradigms where the number of cells is 3,7,15,..., i.e., $2^n - 1$. We note this, but leave it as an unexplored aspect of the system. Total syncretism appears to exist, of course, and we do not exclude it across the board.

4.2 More on the 3-cell space: Pāṇini revisited

Thus far, we have examined only the three minimally valid inventories that generate a three-cell paradigm. In the previous section, we were able to present a complete discussion of all the possible inventories and of the partition sets described by each inventory. There were only 8 possible inventories for the features definable over a two-cell paradigm, and 4 inventories were invalid. But for a three cell space, there are 128 inventories, and numerous sequences to consider.

Table 3 provides a summary of important aspects of the grammar of three-celled paradigms and the models that generate them. In the next paragraphs, we walk through this table in some detail, identifying various properties that are of potential interest. Among these, we note that imposing Pāṇinian ordering — limiting all models to intrinsic rule ordering — turns out to have rather drastic consequences.

Table 3 is divided horizontally into two halves. Each half tabulates all the valid feature inventories, and counts inventories grouped by the number of features they contain (y-axis) × the partition sets that may be generated from them (x-axis). The two halves of the table differ as follows: In the top half, it is assumed that extrinisic order of rules of exponence is permitted, while in the bottom half, we add the additional assumption that only intrinsic (Pāṇinian) ordering is permitted. We discuss the differences below.

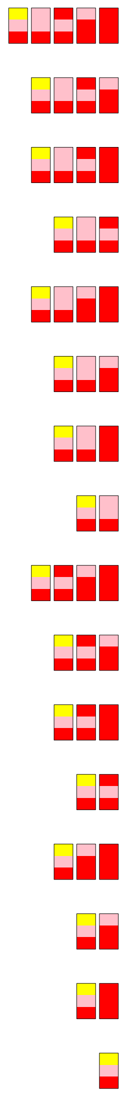
The columns in Table 3 represent possible partition sets of a three-cell paradigm space, using colour instead of letters, as in (25) above: the same colour in two cells indicates the same exponent (syncretism). There are $B_3 = 5$ distinct partitions (the rightmost column) and 16 different subsets of partition that contain the maximally differentiated partition (ABC = red, pink, yellow).

The header of each column represents a distinct partition set, and the number in a given column represents the number of formally distinct (valid) inventories that can in principle generate that set. In the leftmost column of the line "3 features, order", one finds the number 1. Assuming extrinsic rule ordering is allowed, there is exactly one choice of an inventory with three features, from among the 7 possible features, which yields only an ABC partition. We have seen that already; it was the inventory in (3). If that inventory is chosen, from among the 128 possible inventories, then the only partition that can be generated is ABC.

On the same line, the number in the rightmost column is 3. There are (exactly) three distinct choices of feature inventories from each of which all five logically possible inventories can be derived. One such inventory is $f_{110}, f_{101}, f_{111}$, i.e. it is derived from a valid two-feature inventory by adding the total default f_{111}. The other two inventories are also of this type.

This line also shows that there are 3-feature inventories that generate a partition set which excludes ABA. For example, the 5th columnn from the right notes that there are three inventories whose partition sets contain ABC, AAB, ABB, and AAA, but not ABA. One of the three inventories which generate this partition set is $f_{001}, f_{011}, f_{111}$ as we saw above already (the containment inventory). A second possibility is $f_{100}, f_{110}, f_{111}$ (a linear permutation of the previous one). Finally, also the inventory $f_{100}, f_{001}, f_{111}$ generates this partition set. Furthermore, all three inventories exclude *ABA from their corresponding partition set regardless of whether extrinsic ordering is allowed or not.

Bear in mind that the numbers in this table do not count models or sequences, but count inventories. Other than those in the leftmost column, each



2 features, order	–	–	–	–	–	–	1	–	–	–	1	–	1	–	–	–
3 features, order	1	–	2	–	2	–	3	3	2	–	3	3	3	3	1	3
4 features, order	–	–	1	–	1	–	3	2	1	–	3	2	3	2	3	14
5 features, order	–	–	–	–	–	–	1	–	–	–	1	–	1	–	3	15
6 features, order	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	6
7 features, order	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
total with order	1	–	3	–	3	–	8	5	3	–	8	5	8	5	8	39
2 features, Pāṇini	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
3 features, Pāṇini	4	–	4	–	4	–	–	3	4	–	–	3	–	3	1	3
4 features, Pāṇini	–	–	3	–	3	–	1	2	3	–	1	2	1	2	3	14
5 features, Pāṇini	–	–	–	–	–	–	1	–	–	–	1	–	1	–	3	15
6 features, Pāṇini	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	6
7 features, Pāṇini	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
total Pāṇini	7	0	7	0	7	0	2	5	7	0	2	5	2	5	8	39

Table 3 Table showing which 12 of the 16 logically possible three cell partition sets that can be generated by feature systems and by how many.

valid inventory in the table may be contained in multiple models, thus yielding sets of generable partitions. As a concrete example, all of the information in (24) (=(25)) is here coded by the number 1 in the top line, column 11. From the feature inventory in (24), all and only the three partitions at the top of the column (ABC, ACC, AAC) can be generated; moreover, this is the only choice of (two) features which generates that exact partition set (and requires extrinisic rule ordering to do so).

Results

The three minimal valid inventories that we have discussed above are in the top row of the top half of the table. These are the only three valid, two-feature inventories.

One point of interest is that there are four partition sets that are underivable: four columns total to zero (in fact the same four with or without a limitation to Pāṇinian ordering). As the second column shows, there is, for example, no valid inventory (minimal or otherwise) that admits all and only the maximally and minimally differentiated partitions (ABA, AAA). Also excluded are patterns that allow ABC, AAA and exactly one 2:1 grouping (columns 4, 6, 10).

This latter fact is particularly interesting, since the last of these (column 10) is what Bobaljik (2012) finds empirically for suppletion in adjective gradation: ABA and AAB are unattested, but the other patterns are allowed. Our result means that the suppletion pattern of gradation isn't predicted by any variation of the morphological assumptions we consider here – i.e. whether Pāṇini, Minimality or similar condition is assumed. However, Bobaljik also proposes to separate the component accounts of *ABA from *AAB in adjectival gradation, arguing that only *ABA is excluded by the logic of features and syncretism, and proposes an additional, syntactic locality condition to exclude *AAB (see also Bobaljik & Wurmbrand 2013).

Pāṇini

Before leaving the domain of three-cell paradigm spaces, we will consider the effect of one additional restriction, namely the idea that there is no extrinsic ordering of rules, and only Pāṇinian ordering. Each of the three valid, minimal feature inventories makes use of two basic overlapping features, and derives a third by the intersection of those two. We showed above that reordering the rules has the effect of deriving syncretic patterns, in effect, by rendering the intersective feature redundant. The order in (21-a) is equivalent to a system that uses only the two basic features, but not their conjunction.

We may consider imposing Pāṇinian-order-only as a restriction on valid sequences, corresponding to the hypothesis that grammars make use of only intrinsic, but not extrinsic, ordering of rules. Comparing the top and bottom halves of Table 3 allows us to evaluate the effects of this assumption, for three-celled paradigms.

One result which we find interesting is that for 3-cell paradigms, imposing Pāṇinian ordering has no effect on the total number of valid inventories. (This turns out to be different for 4-celled paradigms). We simply note this here, without further comment.

However, comparing the first line of each half of the table shows that imposing Pāṇinian ordering in addition to Minimality is a severe restriction. This constellation of assumptions has the effect that only the maximally distinct partition is describable (the leftmost column in Table 3). All three valid minimal inventories will derive that order and no other. Technically, intrinsic ordering does not restrict the relative order of f_{110} and f_{011}, but since the conjunction will identify the middle cell, the remaining ordering is free (the two are non-distinct).

Since syncretism is abundant in paradigms of all sizes, imposing a Pāṇinian ordering, along with the other assumptions considered above, seems pathologically over-restrictive.

5 Beyond three cells

As we move beyond a three-cell space, the system grows and changes in various ways. Unlike the two- and three-cell spaces, we will not walk through all examples in as much detail, but will call attention to various points and provide some more general discussion in the abstract.

A three-cell paradigm space allows for $2^3 - 1 = 7$ features, and thus $2^7 = 128$ logically possible inventories. We showed in the previous section that Validity reduces that set to 96 and Validity plus Minimality further reduces the number of contenders to 3 feature inventories. By countenancing arbitrary ordering of the rules of exponence (sequences), each of the three minimally valid inventories generates 3 of the $B_3 = 5$ partitions of a 3-member set: none derives the undifferentiated partition. Adding Pāṇini as a restriction reduces that further: from the three minimal valid feature inventories, only one partition can be generated–the maximally differentiated partition.

The four-cell case allows $B_4 = 15$ partitions (paradigms) including the maximally differentiated partition. Logically, $2^{B_4 - 1} = 16\,384$ different valid partition sets exist. If there were no constraints on feature inventories or rules of exponence, this is the number of possible states of the world that we could investigate typologically, to ask which is the actual state. This number is so large that the answer to the question which of these subsets can be derived from how many valid feature systems cannot be visualized in the same way as in table 3. To be able to still compare the three and the four cell results, we report therefore comparable summary statistics for both the three cell case in Table 4 and the four cell case in Table 5. In the three cell case, the values shown in table 4 can be mostly read off from the table in 3. The first column shows the size of the sets of features under consideration, ranging from the minimum number such that there is a valid feature set of that size to the size of the set of all features, i.e. $2^c - 1$ where c is the number of cells. The second column shows the number of subsets of that size whether valid or not, i.e. the binomial coefficient $\binom{2^c-1}{\#\text{features}}$. The next column lists how many of these features sets are valid inventories. This corresponds to the sum over the values of the corresponding row in table 3 – recall that valid feature sets are valid regardless of whether order is extrinsic or intrinsic. The penultimate column indicates how many different ordered partition sets (OPSs) can be derived from the valid feature sets. This corresponds to the number of non-zero entries in the corresponding row of table 3. The last column shows the number of Pāṇinian partition sets derivable from the valid Inventories. The last row of table 3 gives the values for feature sets of any size. The number of inventories and that of valid inventories in the last row are the sum of the entries for feature sets of a specific cardinality because there can't be any overlap. But the paradigm sets can overlap – a paradigm set is derivable from a three feature inventory can in some cases also be derived from a four feature inventory. Therefore the values in the last row of the #PPS and #OPS columns aren't the column sums.

#features	#Inventories	#validInventories	#OPS	#PPS
0	1	0	–	–
1	7	0	–	–
2	21	3	3	1
3	35	29	12	9
4	35	35	11	11
5	21	21	5	5
6	7	7	2	2
7	1	1	1	1
any	128	96	12	12

Table 4 Summary statistics of logically possibly three cell partitions

For the four-cell paradigm space, the picture changes drastically. This is largely because describing a four cell space requires at least three features (as we show below). The table corresponding to table 4 for four cells is shown as table 5. We computed the table using an Apple Macbook Pro Laptop computer (2012 model) running a program in the R programming language (R Core Team 2012) that tests all the feature sets for validity (computer code available). The total computation time amounted to several hours.

#features	#Inventories	#validInventories	#OPS	#PPS
0	1	0	–	–
1	15	0	–	–
2	105	0	–	–
3	455	140	116	47
4	1 365	1 015	317	239
5	3 003	2 793	347	402
6	5 005	4 935	310	420
7	6 435	6 425	240	369
8	6 435	6 435	160	279
9	5 005	5 005	99	193
10	3 003	3 003	54	112
11	1 365	1 365	25	61
12	455	455	12	24
13	105	105	6	6
14	15	15	2	2
15	1	1	1	1
any	32 768	31 692	361	463

Table 5 Summary statistics of logically possibly four cell partitions

The results for the four-cell systems differ noticeably from the three-cell case in terms of the effect of the Panini assumption and the overall restrictiveness. The biggest difference is that in the three cell case the minimal inventories contain 2 features, but for four cells a least 3 features are required. Validity and Minimality are about equally restrictive, but in relation to the 256 times greater set of feature sets. This leaves 140 minimal valid feature inventories. Unlike the three-cell space, where adding Pāṇini eliminated the possibility for variation, in the four-cell space, Pāṇinian sequences allow 47 distinct possible partition sets, each of which generates between 1 and 7 of the $B_4 = 15$ distinct partitions of a 4-celled space. None generate more than 7 partitions. The occurrence of the number 47 – a prime number – in the table is, we think, noteworthy: It might indicate that even though the concept of a Pāṇinian sequence is a straightforward mathematical concept, the concept leads to some irreducible formal complexity. We include a display of these 47 paradigm sets as an appendix.

A second main difference concerns the relation between the order-sensitive partition sets (OPSs) and Pāṇinian ones (PPSs) from (13). In the three cell case, the OPS and PPS were overall the same. But with four cells, sequences with extrinsic order and Pāṇinian sequences predict different restrictions on the paradigm sets. As shown by the last line of table 5, the Pāṇini assumption allows slightly more variation – 463 paradigm sets are generable with Pāṇinian sequences, while only 361 are compatible with extrinsically ordered systems.[10] We furthermore computed the amount of overlap between the OPSs and PPSs. We found that the union of OPSs and PPS contains 557 partition sets, and therefore 267 are both OPS and PPS, 94 are only OPSs, and 196 are only PPSs. Some examples are shown in table 6: the first three rows show three of the 267 partition sets that are both OPSs and PPSs. The second block of three rows shows three that are only PPSs, and the third block shows two that are only OPSs. The last block illustrates with two examples the 15 827 partition sets that cannot be generated by a feature based morphological system unless additional restrictions are introduced.

Overall feature-based analyses are surprisingly restrictive in the four-cell case. Of the 16 384 logically possible subsets of paradigms, only fewer than 3.5% are compatible with a feature based approach, 15 827 logically conceivable typological states are ruled out. Pāṇini's assumption is even slightly more restrictive, allowing less than 3% of all possible sets of partitions. This is surprising since the feature assumption and also the Pāṇini assumption only ruled out 4 of 16 paradigm sets in the 3 cell cases, i.e. 75% of the paradigm sets are compatible. Even assuming minimal Pāṇini-systems turns out to be

10 It may seem counter-intuitive that systems that allow any order of rules generate less variation than systems which are limited only to Pāṇinian rule orders. In terms of individual partitions, since Pāṇinian orders are a (proper) subset of the orders that are statable as extrinsic orders, Pāṇinian systems will not generate fewer partitions than those with unrestricted ordering. But what we count here is not the partitions that are generated, but the partition sets. In general, allowing extrinsic order will mean that the partition set generable from a given inventory will be a superset of the set generable from the same inventory with only Pāṇinian order (as mentioned in note 6). The numbers above reflect the effect that there is more convergence among the larger sets: with unrestricted rule ordering, more distinct inventories converge on the same partition sets than in the case of restricted rule ordering, overall.

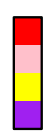
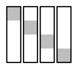
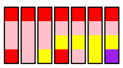
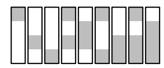
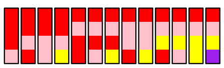
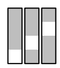
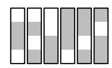
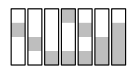
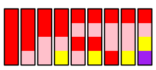
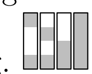
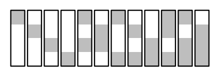
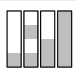
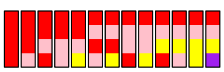
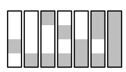
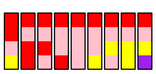
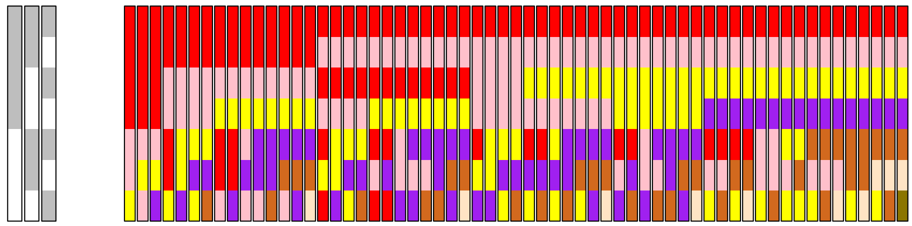
partition set	features (with order)	features (Panini)
	1: 	57, e.g. 
	88, e.g. 	90, e.g. 
	1024, e.g. 	128, e.g. 
	0	32, e.g. 
	0	2, e.g. 
	0	128, e.g. 
	2, e.g. 	0
	12, e.g. 	0
	0	0
	0	0

Table 6 Examples of four-cell partition sets with the number of ordered and Panini analyses and example feature systems

more restrictive in the four cell than in the three cell case, allowing less than 0.3% of the paradigm sets with four cells compared to 1 of 16 with three cells. The numerical comparisons need to be taken into account when evaluating feature based morphological theories against typological data since the more restrictive a constraint is the less likely it is that is satisfied by chance.

In many ways, the constraints imposed by feature systems on universal grammar are at this point difficult to grasp concretely, though future formal work may find a way to do so. We explore therefore a four cell version of the *ABA constraint, *ABBA (see Caha 2009; Smith et al. 2016 for empirically-based proposals that exclude this pattern). There are two other permutations of the ABBA partition, ABAB and AABB. We investigated therefore which partition sets containing all three of ABBA, ABAB and AABB, as well as ABCD, can be generated from feature and Pāṇini systems. There are 2^{11}, i.e. 2 048 such feature sets, but only 41 of them can be generated by feature systems. Of those, 21 can be generated with either ordered or Pāṇinian systems, and a further 10 each with either only an ordered or only a Pāṇinian system. The examples in the third and fifth row of table 6 illustrates two such systems. The system in row three predicts 13 partitions and can be generated from only three features if order is allowed. But if Pāṇini's restriction is obeyed, at least six features are necessary to generate this system. Row 5 shows an ABBA-system with the smallest number of features required to be generated in a Pāṇinian way, namely 4. So in the Pāṇini case, ABBA-systems require more than the minimal number of features as we also found for ABA-systems.

To conclude this section, consider briefly the seven cell paradigm. Because both 3 and 7 have $2^n - 1$ cells for some n, the structures of the two paradigms are similar. The minimal number of features of a valid inventory for a seven cell paradigms is three. Any set of three features f_a, f_b, and f_c where each basic feature contains 4 cells, each conjunction of two features contains 2, and the conjunction of all three features contains 1 cell is valid. There are $7! = 5040$ such systems, and all have the same structure except for permutations of the cells. One system is shown in (26) along with its order-sensitive partition set (OPS) – like in teh 3-celled paradigm space, only the maximally differentiated paradigm is possible with Pāṇini sequences. It is noticeable that of the 61 generable partitions in the OPS none contains only A and B – only paradigms with at least a three-way partition are possible. Since all OPSs are identical except for cell-permutations, we can derive in this way a *AB constraint for seven cell paradigm spaces. If there are such spaces, we make at least this typological prediction.



6 Discussion

6.1 *ABA - empirical considerations

Coming up out of the heady sea of numbers for air, we are now at a point to step back and ask whether the results of our investigation of the formal combinatorics of features has any bearing on the actual *ABA generalizations discussed in the literature. Our tentative conclusion is that some domains where a *ABA generalization is observed do not seem to conform to the profile of a Minimal Valid Inventory, while for others, the situation is less clear, and the Minimal Valid Inventory, with overlapping features, rather than containments, seems to us to be a direction worth pursuing.

We opened this article with reference to the *ABA generalization in adjectival gradation, investigated extensively in Bobaljik (2012). We see no reason from the discussion here to think that it would be profitable to reanalyze that as arising from a Minimally Valid Feature Inventory. Doing so would invoke two privative features, one shared by the positive and comparative grade (but not the superlative), and another shared by the comparative and superlative, but not the positive. There is, however, fairly extensive evidence independent of patterns of suppletion for a containment relation in adjectival gradation: the superlative transparently contains the comparative in many languages. Some examples are given here (from Bobaljik 2012:31):

(26)

	POS	CMPR	SPRL	
a. Persian:	**kam**	**kam**-tar	**kam**-tar-in	‘little’
b. Cimbrian:	**šüa**	**šüan**-ar	**šüan**-ar-ste	‘pretty’
c. Czech:	**mlad**-ý	**mlad**-ší	nej-**mlad**-ší	‘young’
d. Hungarian:	**nagy**	**nagy**-obb	leg-**nagy**-obb	‘big’
e. Latvian:	**zil**-ais	**zil**-âk-ais	vis-**zil**-âk-ais	‘blue’
f. Ubykh:	**nüs°ə**	ç’a-**nüs°ə**	a-ç’a-**nüs°ə**	‘pretty’

In addition, it is not at all obvious that it makes sense to consider adjectival degrees as grammatical features, in the way that, for example, classificatory elements such as gender are.

On the other hand, there are other domains in which *ABA generalizations have been observed, where there is less independent reason to think that the constituent elements are arranged in a containment relation.

One such domain, perhaps, is person. Vanden Wyngaerd (2016) sees an *ABA generalization in (plural) independent pronouns. Building on prior cross-linguistic investigations (Cysouw 2003; Baerman et al. 2005), he observes that there are languages where first and second (plural) pronouns are syncretic, contrasting with the third person (such as Slave, in (27), from Cysouw 2003:124), and there are languages where second and third (plural) are syncretic, contrasting to the first person (as in the Nez Perce 'unmarked' pronouns in (28), Cysouw 2003), but virtually no good examples of syncretism of first and third person, contrasted with second.[11]

(27)

	SG	PL
1	sį	naxį
2	nį	naxį
3	ʔedį	ʔegedį

(28)

	SG	PL
1	'íin	núun
2	'íim	'imé
3	'ipí	'imé

Vanden Wyngaerd (2016) argues for a containment relation among the features that define person, as in the following:[12]

(29) a. 1^{st}: [[[PERSON] PARTICIPANT] AUTHOR]
b. 2^{nd}: [[PERSON] PARTICIPANT]
c. 3^{rd}: [PERSON]

In our terms, this is (a linear permutation of) the inventory in (5): $f_{100}, f_{110}, f_{111}$ and its properties are well understood. However, there are few, if any, languages in which such a decomposition of pronouns is transparently manifest in surface forms. As we have seen above, this inventory is valid, but non-minimal.

11 In bound person marking (agreement) more patterns are attested, though of varying frequency (Cysouw 2003, 2010).

12 This is one of the current prominent views about the decomposition of person features in the literature; see for example: Sauerland (2008); Zeijlstra (2015). For contrasting views, see Bobaljik (2008); Harbour (2016).

A Minimal Valid inventory would be one that composes the three persons out of two privative features: f_{110} corresponding to the feature 'participant', and f_{011}, which is in essence the privative feature 'non-author'. On this alternative analysis, first and third person pronouns cannot be syncretic, excluding the second person, since they share no feature. Hence *ABA.

In work in progress (see Sauerland & Bobaljik 2013) we are exploring the typology of syncretism in person feature systems more broadly, drawing on the extensive data in Cysouw (2003), to determine what feature system has the maximum likelihood of underlying the observed partition sets, not just in plural pronouns, but in the full range of person marking systems, including clusivity distinctions. We may wager that if we are right to suspect a Minimally Valid Inventory at work in the patterns of syncretism in the free-standing pronouns, then we will see that emerge as well in the larger study.

Before closing, we note as well that *ABA generalizations have also been noted in verbal inflection (Wiese 2008; Starke 2009), case (Caha 2009; Smith et al. 2016), and number (Smith et al. 2016). Of these, case is another domain in which there is minimal independent morphological evidence for containment relations, at least among 'core' cases.[13]

Pavel Caha (personal communication and this volume) calls our attention to at least one sub-part of the case hierarchy which appears to reflect the kind of feature structure we would expect on the approach taken here. Blansitt (1988) surveys the marking of the following four functions across the world's languages: direct object, dative (recipient), allative (goal of motion), and location. Blansitt notes a generalization, exceptionless in transitive clauses, whereby no two functions are marked identically unless all intervening functions in the order just given are also marked identically. In other words, a *AB(B)A generalization. One way to approach this, following Caha (2009) (but see also Caha & Pantcheva 2012), would be to assume that there is a monotonic containment relationship among the features (we consider the last three for ease of exposition):

(30) a. f_{111} = dative
b. f_{011} = allative
c. f_{001} = locative

An alternative, following the approach laid out here, would be the minimal valid inventory in (31):

13 Caha (2009); Smith et al. (2016) report some examples of, e.g., dative built on accusative, etc., but these are surprisingly rare, in contrast to, e.g., what we find with adjectival gradation. For spatial/locative cases, there is a much richer amount of transparent embedding; see Comrie & Polinsky (1998); Radkevich (2010); Pantcheva (2011).

(31) a. f_{110} = dative
b. f_{011} = locative

From this inventory, the allative can be described as the intersection of the other two cases. As Caha notes, Blansitt offers at least one language that seems to transparently reflect (31) rather than (30). Tigrinya prepositions include *ne* which marks dative (and some objects, presumably an instance of differential object marking, which quite commonly uses the dative, Bossong 1985) and locative *ab*. The allative is marked by the conjunction of the two: *nab* < *ne ab*. This is also broadly consistent with the results of Radkevich (2010) who found no evidence of a simple, monotonic transparent relationship among local cases as (30) might predict (although her survey also finds cases of portmanteaus and internally complex case morphology that are equally hard to reconcile with (31)).

We note also that as paradigms grow, the type of representation entertained here readily accommodates multi-dimensional syncretism, a prima facie challenge for theoretical approaches, such as Nanosyntax, which adopt a universal total (containment) ordering among features (see Caha & Pantcheva 2012 for ideas on how to extend the Nanosyntax model to accommodate this.) We have throughout represented paradigms as one-dimensional lists, as in (32), although one often finds four-celled paradigms presented as a 2 × 2 matrix, encoded as two binary features, as in (33).

(32) <A,B,C,D>

(33)

	$-\alpha$	$+\alpha$
$-\beta$	A	B
$+\beta$	C	D

Translation is straightforward: the feature $-\alpha$ in (33) is encoded relative to the list in (32) in our terms as f_{1010}, $+\beta$ as f_{0011}, etc. But horizontal and vertical syncretisms have no a priori special status — we can just as readily define a feature f_{1001} which picks out cells A and D, a diagonal syncretism in (33). For us, this flexibility is an advantage, since it allows us to take any existing partition set and probe what the optimal underlying feature inventory might be, given any combination of assumptions such as binarity, Pāṇinian order, Minimality etc. Rather than setting the features ahead of time, we can in this way discover whether features should be binary or not.

Consider in this light the first line of the third block in Table 6. The partition set contains 9 of the $B_4 = 15$ logically possible partitions of the four cell space. If we map the lists in the partition set to a binary table as in

(33), we may observe that this partition set contains partitions corresponding to horizontal and vertical syncretisms (the third and fifth partition sets, respectively), but no diagonal syncretisms. If this partition set is what we observe typologically (i.e., all attested paradigms correspond to one of these 9 partition sets), does this indicate that the inventory should contain binary features? Evidently not, since the feature inventory that generates this partition, given in the second column, does not contain binary features. It is a minimally valid inventory, containing the three features in (34) (and thus allowing the intersection of the first two in the Rules of Exponence):

(34) a. f_{0101}
 b. f_{0011}
 c. f_{1111}
 d. $(f_{0101} \cap f_{0011} = f_{0001})$

In Sauerland & Bobaljik (2013), we note that four the four-cell paradigm space corresponding to the first person (inclusive vs. exclusive × singular vs. plural), 9 of the 15 possibilities are indeed attested (Cysouw 2003), and a feature inventory along these lines provides the optimal analysis of the space of typologically attested possibilities.

Without probing deeper, we hope to have shown that the derivation of *ABA generalizations entertained here may indeed get off the ground in some domains, leaving for future work the fuller empirical investigation of this approach.

Why Minimality?

Finally, returning to the question we raised at the outset, we may step back even further and ask why UG might have the types of constraints it does. We are obviously far from an answer, but can add a few, very tentative remarks here.

To this point, we have assumed that it is reasonable to think that UG feature inventories respect a condition of Minimality, and have shown how this assumption restricts the hypothesis space to be considered in determining the actual feature inventory corresponding to paradigms of a given size. Minimality has a somewhat different flavour than some of the other restrictive assumptions we have entertained. In principle, one could think of this from a different perspective. Rather than imposing a condition of Minimality on inventories, one could imagine instead that the features are whatever they are, but that UG shows maximal use of the features it has. For a domain with two features, UG generates in principle a three-celled space: each feature on its own, plus their intersection. This builds in the assumption of minimality - and thus means that all true three-celled paradigms are those projected from the two-feature inventories, yielding the *ABA prediction (up to linear permutation).

This alternative (maximal use of minimal resources), implies that there should be no four, or five-celled paradigms. If there are two features (in a given domain) then the maximal paradigm in that domain will have three cells. If there are three features, then the paradigms generated will have 7 cells. The appearance of a four-celled paradigm in some domain then necessarily involves syncretism.

Acknowledgments

For feedback while developing the ideas presented here, we thank audiences at Universität Leipzig, Goethe-Universität Frankfurt, and the Leibniz-Zentrum Allgemeine Sprachwissenschaft. We are particularly grateful to Kazuko Yatsushiro and to Tom Green for his contributions to our understanding of the combinatorics. This work has been financially supported in part by the NSF (grant BCS-0616339), the Alexander von Humboldt-Stiftung, and by the Bundesministerium für Bildung und Forschung (BMBF) (Grant 01UG1411).

References

Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The syntax-morphology interface: a study of syncretism*. Cambridge: Cambridge University Press.

Blansitt, Edward L. 1988. Datives and allatives. In Michael Hammond, Edith A Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology*, 173–191. Amsterdam: Benjamins.

Bobaljik, Jonathan David. 2002. Syncretism without paradigms: Remarks on williams 1981, 1994. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2001*, 53–85. Dordrecht: Kluwer.

Bobaljik, Jonathan David. 2008. Missing persons: A case study in morphological universals. *The Linguistic Review* 25(1-2). 203–230.

Bobaljik, Jonathan David. 2012. *Universals in comparative morphology: Suppletion, superlatives, and the structure of words*. Cambridge: MIT Press.

Bobaljik, Jonathan David & Susi Wurmbrand. 2013. Suspension across domains. In Ora Matushansky & Alec Marantz (eds.), *Distributed Morphology today: Morphemes for Morris Halle*, 185–198. Cambridge: MIT Press.

Bossong, Georg. 1985. *Empirische universalienforschung. differentielle objektmarkierung in den neuiranischen sprachen*. Tübingen: Narr.

Caha, Pavel. 2009. *The nanosyntax of case*. Ph.D. thesis, CASTL Tromsø.

Caha, Pavel & Marina Pantcheva. 2012. Datives crosslinguistically. Unpublished handout CASTL.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2. 113–124.

Comrie, Bernard & Maria Polinsky. 1998. The great dagestanian case hoax. In Anna Siewierska & Jae Jung Song (eds.), *Case, typology, and grammar*, 95–114. Amsterdam: Benjamins.

Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford, UK: Oxford University Press.

Cysouw, Michael. 2010. On the probability distribution of typological frequencies. In *The mathematics of language*, 29–35. Springer.

Harbour, Daniel. 2008. On homophony and methodology in morphology. *Morphology* 18(1). 75–92.

Harbour, Daniel. 2016. *Impossible persons*. Cambridge: MIT Press.

Jakobson, Roman. 1936/1971. Beitrag zur allgemeinen kasuslehre. gesamtbedeutungen der russischen kasus. In *Selected writings*, vol. 2, 23–71. The Hague: Mouton.

Kiparsky, Paul. 1973. ”Elsewhere” in phonology. In *A festschrift for Morris Halle*, 93–106. Holt, Rinehart and Winston.

Kiparsky, Paul. 1979. *Pāṇini as a variationist*. MIT Press.

Pantcheva, Marina. 2011. *Decomposing path: The nanosyntax of spatial expressions*. Ph.D. thesis, Universitetet i Tromsø.

R Core Team. 2012. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radkevich, Nina. 2010. *On location: The structure of case and adpositions*. Ph.D. thesis, University of Connecticut.

Sauerland, Uli. 2008. On the semantic markedness of ϕ-features. In David Adger, Susana Béjar & Daniel Harbour (eds.), *Phi theory: Phi features across interfaces and modules*. Oxford: Oxford University Press.

Sauerland, Uli & Jonathan D. Bobaljik. 2013. Syncretism distribution modeling: Accidental homophony as a random event. In Nobu Goto, Koichi Otaki, Atsushi Sato & Kensuke Takita (eds.), *Proceedings of GLOW in Asia*. Tsu, Japan: University of Mie.

Smith, Peter, Beata Moskal, Ting Xu, Jungmin Kang & Jonathan David Bobaljik. 2016. Case and number suppletion in pronouns. Manuscript Goethe Universität Frankfurt, University of Connecticut, and Syracuse University.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1). 1–6.

Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambrdige: Cambridge University Press.

Vanden Wyngaerd, Guido. 2016. The feature structure of pronouns: a probe into multidimensional paradigms. Unpublished manuscript CRISSP.

Wiese, Bernd. 2008. Form and function of verb ablaut in contemporary standard german. In Robin Sackmann (ed.), *Explorations in integrational linguistics*. Amsterdam: John Benjamins.

Zeijlstra, Hedde. 2015. Let's talk about you and me. *Journal of Linguistics* 51. 465–500.

Appendix: The 47 Four Cell Pāṇinian Partition Sets (PPPs)