ICLC Twentieth Anniversary Volume (1986)

IMPOSED VERSUS INHERENT FEATURE SPECIFICATIONS, AND OTHER MULTIPLE FEATURE MARKINGS*

ARNOLD M. ZWICKY

Ohio State University and Stanford University

1. Introduction

This paper is intended as a contribution to the theory of features in generative grammar. Since generalized phrase structure grammar (GPSG) has the most developed theory of this sort now available--and the only one that has been fully formalized--my remarks will be framed as proposals for amending GPSG, but they should be understood as having wider applicability.

The feature theory formalized in Gazdar, Klein, Pullum, and Sag 1985 (hereafter GKPS) permits a category to have only a single value for any given feature. An NP can have only one value for CASE, one value for PERS[on], one value for NUM[ber], one value for GEND[er], and so on, not to mention only one value for BAR and for each of such category features as N, V, and MAJOR. I will suggest, largely by citing familiar facts, that this simple and attractive proposal is inadequate, that multiple feature specifications must be allowed--in particular, that specifications imposed on a constituent by rules of government and agreement (which I will refer to as IMPOSED SPECIFICATIONS) must be kept separate from specifications that are not determined in this way (which I will refer to as INHERENT specifications), and that of the specifications imposed by agreement on a constituent, several must be distinguished according to the types of triggers that impose them (and that similar distinctions perhaps must be made among specifications imposed by government).

2. Possessive Marking

The need for multiple feature marking can be seen quite clearly in the possessive constructions of nearly any language with a moderate amount of inflectional morphology. I will give examples from German and Turkish, the first illustrating agreement of a possessor (PR) with a possessed constituent (PD), the second the reverse.[1]

2.1 German Possessives

The personal pronouns in German have several case forms. In the 3 SG MASC, the forms are NOM *er*, ACC *ihn*, DAT *ihm*, GEN *sein*, the last of

these patterning with other prenominal GEN forms of definite NPs, like *Jakob-s* 'Jacob's', and like them representing PR in combination with PD: *Jakobs Buch* 'Jacob's book', *sein Buch* 'his book'. The 3 SG FEM forms are NON and ACC *sie*, DAT and GEN *ihr*.

These case forms are indeclinable, except for GEN forms like *sein* and *ihr*, which show agreement with PD in case, gender, and number and have patterns of suffixation just like the indefinite article *ein*:

	'his father'	'her father'	'his mother'	'her mother'
NOM	sein Vater	ihr Vater	sein-e Mutter	ihr-e Mutter
ACC	sein-en Vater	ihr-en Vater	sein-e Mutter	ihr-e Mutter
DAT	sein-em Vater	ihr-em Vater	sein-er Mutter	ihr-er Mutter
GEN	sein-es Vaters	ihr-es Vaters	sein-er Mutter	ihr-er Mutter

All of these PR forms illustrate multiple feature marking. Consider, for instance, a direct object NP *seine Mutter* 'his mother'. The base *sein* must be represented as having the specifications CASE:GEN GEND:MASC NUM:SG; CASE:NOM GEND:MASC NUM:SG would be *er*, CASE:GEN GEND: FEM NUM:SG would be *ihr*, and CASE:GEN GEND:MASC NUM:PL would be *ihr*. But the full form must be represented as having the specifications CASE:ACC GEND:FEM NUM:SG in order to predict the inflectional material. The inherent specifications of a PR pronoun must be distinguished from the specifications it bears by agreement with PD.

2.2 Turkish Possessives

The construction under consideration here[2] is a subtype of a more general construction consisting of PR followed by PD. PR has suffixes of the GEN case appropriate for its person and number (1 SG *-im*, 2 SG *-in*, 3 SG *-in*, etc.), and PD has suffixes (variants of the GEN suffixes just mentioned) indicating the person and number of PR: *Mehmed-in el-i* 'Mehmet's hand'. A pronominal PR like *ben-im* in *ben-im el-im* 'my hand' is omissible, information about its person and number being available in the suffixes of PD, as in the paradigm below:

el-im 'my hand'	el-im-iz 'our hand'
el-in 'your (SG) hand'	el-in-iz 'your (PL) hand'
el-ler-im 'my hands'	el-ler-im-iz 'our hands'
el-ler-in 'your (SG) hands'	el-ler-in-iz 'your (PL) hands'

To forms like these correspond further forms with case suffixes: to *ellerimiz* 'our hands', ACC *ellerimiz-i*, DAT *ellerimiz-e*, LOC *ellerimiz-de*, even GEN *ellerimiz-in*. All of the forms above illustrate multiple feature marking--in *el-im* the base *el* is 3 SG and the suffix *-im* is 1 SG, for instance--but the last form, *ellerimizin*, is a real marvel of multiple feature marking: the base *el* represents the specification PERS:3 and the suffix *-ler* the specification NUM:PL, and the two together indicate no value of CASE for PD; the suffix *-im* represents the specification PERS:1 and the suffix *-iz* the specification NUM:PL, and the two together represent the specification CASE:GEN for PR; and the suffix *-in* represents the specification CASE:GEN for PD. Spread across this form, then, are separate indications of the values of CASE, PERS, and NUM for PD and for PR.

There are a number of interesting analytic issues here: how to represent the fact that in German PR picks up specifications of PD, whereas in Turkish PD picks up specifications of PR; whether forms like *elim* pick up their PR specifications by agreement with (an often empty) PD, or whether these PR specifications are represented on the NP node dominating PD and so are inherited by PD as the head of this NP; and so on. My immediate focus, however, is on the prior observation that the inherent specifications of PD must be distinguished formally from the specifications that are imposed on PD by agreement with PR, because the two sets are separately realized in the inflectional apparatus of PD.

3. Verbal Number

Another set of instances of multiple feature marking, quite similar in character to the possessive examples of section 2, has been discussed by Durie 1986 in connection with the phenomenon of VERBAL NUMBER--that is, number as an inherent categorization of verbs--as manifested in a diverse collection of the world's languages.

According to Durie, the inherent category of verbal number corresponds to a property of events, namely their multiplicity, rather than of things, and it is in principle independent of number imposed by grammatical agreement. In many of Durie's examples verb bases coding different verbal numbers are suppletively related to one another, as are the pronoun bases coding different inherent numbers in German (SG *sein* versus PL *ihr*): Huichol 'kill', with SG base *mieni* versus PL *qiini*, the first used for single events of killing (whether with a SG subject or, as below, with a PL one) and the second for multiple events of killing (whether with a SG subject, as below, or with a PL one).

me - neci - mieni	ne - wa - qiini
3 PL SU 1 SG DO SG	1 SG SU 3 PL DO PL
'They are killing me.'	'I am killing them.'

I am not proposing to justify Durie's analysis of these examples (and others) as involving inherent number. Instead, I merely point out that if any of his analyses are correct, then in some languages inherent values of the feature NUM on verbs must be distinguished from values of NUM imposed on verbs by agreement.

4. Argument Marking on Verbs

The Huichol examples above also illustrate a very common feature of languages with complex inflectional morphology: marking of verbs with respect to more than one argument. The Huichol verbs show separate marking for the person and number of subject (SU) and for the person and number of the direct object (DO). Both the SU and the DO features are imposed on verbs, and it seems clear that some way must be found to distinguish formally among subsets of imposed specifications.

In the brief subsections that follow I enumerate some phenomena that might suggest conditions on an adequate general theory of multiple agreement marking: specifications combined from several arguments, more than two choices of argument, marking of more than two arguments in a single form, arguments that aren't subcategorisands, and second-order agreement. The data alluded to here are not equally clear in their interpretation; in some cases I mean only to suggest that they be examined in the light of the formal issues I will be bringing up.

4.1 Combining Specifications From Several Arguments

A principle governing the selection of inflectional affixes in a language can refer to more than one argument at once. This is certainly true for the agreement prefixes of transitive verbs in the Algonquian languages. The general scheme there is for a verb to have a *k* prefix if either its SU or its DO involves a 2nd person, an *n* prefix if either its SU or its DO involves a 1st person, and a *w* or zero prefix otherwise; see Anderson 1977 and Zwicky 1977 for discussion and references.

4.2 More Than Two Choices of Argument.

Argument marking is by no means restricted to the two arguments SU and DO, though its limits are unclear. In the Indic language Maithilī, according to Stump's (1980:5) summary of Jhā's 1958 description, verbs 'agree in person and honorific grade not only with their subject, but with any one of their oblique objects, or with a genitive noun phrase modifying one of these'. Maithilī verb inflections typically combine specifications, but from only two arguments: SU and any other argument on a list that includes not only DO but also obliques and PR NPs associated with any one of these arguments, SU included. And according to Anderson (1985:196), 'Chinook, and the languages of the Northwest Caucasian family, are examples of languages in which up to four noun

phrases can be represented in the verb.' The additional arguments that may on occasion be so represented include indirect objects, benefactives, instrumentals, and motional complements.

4.3 Marking of More Than Two Arguments in a Single Form

Unlike Maithilī, where only two arguments are marked in any one verb form, the examples Anderson cites are supposed to illustrate the marking of up to four arguments in a single form.

4.4 Arguments That Aren't Subcategorisands

Insofar as instrumentals, benefactives, and the like can trigger verb agreement, not all marked objects are NPs that subcategorize verbs. Such data, if verified by further investigations, would frustrate general proposals, cited by GKPS (107:fn 20), to collapse the list of agreeing NPs with the list of subcategorizing NPs.[3]

4.5 Second-order Agreement

The Maithilī data also illustrate agreement at one remove, with (PR) constituents of arguments, as in *tohar bap aelthunh* 'Your father came' (Stump 1980:6), where the verb form *aelthunh* indicates both honorific 3rd person (showing agreement with the SU argument *bap* 'father') and nonhonorific 2nd person (showing agreement with the PR *tohar* of the SU).

4.6 Inverse Forms

Anderson (1977 and elsewhere) has pointed out that certain so-called 'inverse forms' of verbs, as in the Algonquian languages, might have important consequences for a theory of morphosyntactic feature representation. Such forms, Anderson claims, do not involve altering grammatical relations in syntactic structure, though they do have some of the inflectional apparatus that would be expected if SU and DO were exchanged.

5. Issues of Formalization

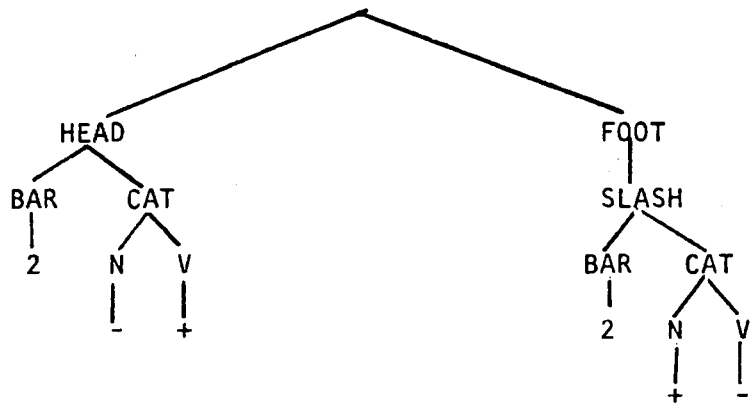
I now step back from data illustrating multiple feature specifications to consider how such specifications might be represented formally. Two related phenomena--feature subsets and category-valued features--have received attention in the GPSG literature, and I will treat multiple feature specifications in the light of the formalisms that have been provided for these related phenomena. I will begin by making the minimal assumption that grammatical categories correspond to sets of feature specifications; the question will be what sort of further structure must be imposed upon such sets.

Consider a set of feature specifications, for instance the category {BAR:2, N:+, V:-, WH:+, CASE:GEN, NUM:SG}. We will frequently have need to refer to FEATURE SUBSETS, to distinguished subsets of feature specifications within this category--for instance, the subset of specifications for syntactic category features {N:+, V:-}, the subset of specifications for head features {BAR:2, N:+, V:-, CASE:GEN, NUM:SG}, and the subset of specifications for foot features {WH:+}.

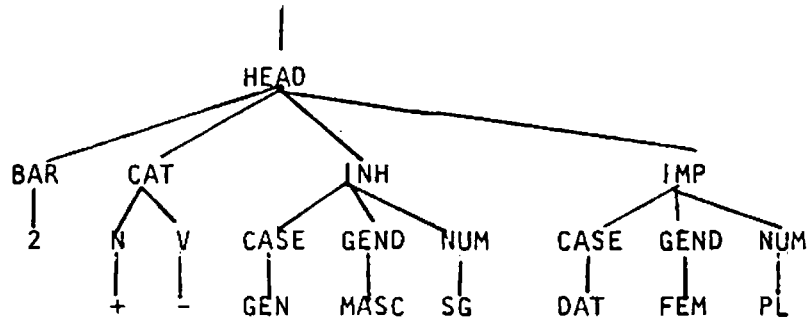
A conceptually different situation is presented by CATEGORY-VALUED FEATURES, features that take grammatical categories as their values. The GPSG feature SLASH is an example; the expression 'VP/NP' refers to a VP missing an NP, that is, to a category with the specifications BAR:2, N:-, and V:+, plus a specification for the feature SLASH that picks out the missing category {BAR:2, N:+, V:-}.

5.1 Tree Formalism

Gazdar and Pullum 1982 (hereafter GP) proposed a formalism that can be used to represent both feature subsets and category-valued features: an unordered-tree formalism in which both names of feature subsets and names of category-valued features can appear as node labels, as in the following representation for VP/NP.



Multiple feature marking of the sort I have illustrated could be accommodated within this formalism by the addition of feature subset names INH[erent] and IMP[osed], so that a masculine singular PR NP in agreement with a dative feminine plural PD N^1 would be represented along the following lines:



GP's actual treatment of agreement put all features involved in agreement, for both the controller and the controllee, under the label AGR. But if the tree formalism is to be extended to PR-PD agreement, then the two sets of specifications INH and IMP must be formally distinguished in some way.

5.2 Set-theoretic Formalism

The GP formalism is versatile, but it does not easily accommodate overlapping feature subsets. If, for instance, we want SLASH to be classified both as a head feature and a foot feature--as GKPS in fact do--then we are obliged to specify values for SLASH separately under the labels HEAD and FOOT, AND to stipulate that these values cannot vary independently (as the values for INH and IMP do in the illustration just given) but must instead always be identical.

Rather than complicate the GP scheme to allow overlapping feature subsets, GKPS opt instead for a straightforward set-theoretic treatment of grammatical categories, with HEAD, FOOT, and the like interpreted as names for sets of specifications for certain features. A specification for SLASH, then, belongs to both the HEAD and the FOOT sets. Unadorned, this framework has no place for multiple feature markings, however; each feature can have only one value in any given category.

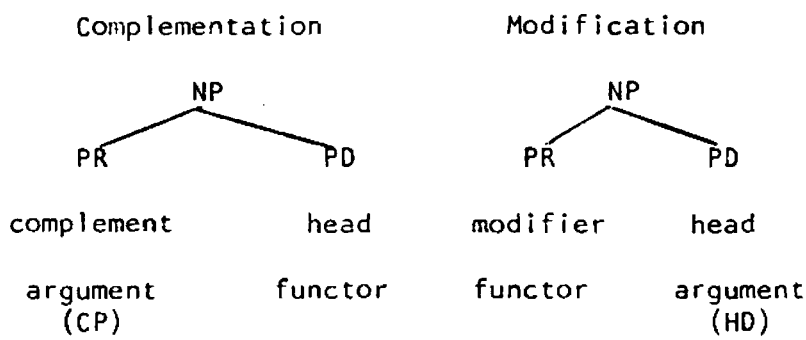
5.3 Exploiting Category-valued Features

GKPS also adopt category-valued features, treating them as an unavoidable complication in their framework. This fact is important in the present context, since GKPS posit a category-valued feature AGR associated with controllers, and this feature distinguishes specifications imposed by agreement from those not so imposed.

Let me illustrate the point with possessives. Suppose that one possible branching of NP is into NP and N^1, these being interpreted as PR modifier and PD head, respectively. Then the GKPS version of the Control Agreement Principle will (in essence) require that the value of the feature AGR for the NP be identical to the set of HEAD specifications

for the N^1. If the N^1 has the HEAD specifications {BAR:1, N:+, V:-, CASE:DAT, GEND:FEM, NUM:PL}, then the NP must have among its HEAD specifications AGR:{BAR:1, N:+, V:-, CASE:DAT, GEND:FEM, NUM:PL}; the NP might have, say, the head specifications in {BAR:2, N:+, V:-, CASE:GEN, GEND:MASC, NUM:SG, AGR:{BAR:1, N:+, V:-, CASE:DAT, GEND:FEM, NUM:PL}}. Realization rules can then refer to CASE:*x* that belongs to a value of AGR--that is, they can refer to a value of CASE imposed by agreement--as distinct from an inherent specification CASE:*y*, which is not a member of a value of AGR. The specifications are, in effect, layered, with the inherent ones being 'parallel with' AGR (being co-members with AGR in their category), the imposed ones being 'inside' AGR (being members of a value of AGR).

In possessive agreement of the German sort, a modifier PR agrees with a head PD, just like other modifiers; the German possessives are in fact declined in much the same way as determiners and adjectives. PD is the argument, PR the functor here, and I will say that PD bears the grammatical relation HD (head) to PR in such a situation. In possessive agreement of the Turkish sort, a complement PR controls agreement on a head PD, just as in subject-verb agreement; indeed these agreement markers on nouns are often identical to subject agreement markers on verbs.[4] PD is the functor, PR the argument here, and I will say that PR bears the grammatical relation CP (complement) to PD in such a situation. Schematically:



I provide no account here of how PRs and PDs can stand in different grammatical relations to one another in different languages, noting only that category-valued features will allow imposed specifications to be distinguished from inherent ones in either situation.

This analysis might be extended to some instances of second-order agreement, assuming that AGR is itself a HEAD feature. Consider a PD SU, of the Turkish type, in a language with subject-verb agreement. The SU NP will share its HEAD feature specifications--including the value of AGR as well as, say, GEN:a_1, PERS:b_1, NUM:c_1--with its head NP (by the Head Feature Convention); this value of AGR is identical to the HEAD feature set of the PR NP (among them, say, GEND:a_2, PERS:b_2,

NUM:c_2). Then the Control Agreement Principle requires that the value of AGR for the VP be identical to the set of HEAD feature specifications for the SU NP, with the result that this VP (and, by the HFC, its head V as well) has three layers of specifications. Schematically:

$$\{N{:}-,\ V{:}+,\ldots,\ AGR{:}o,\ldots\}$$
$$\{GEND{:}a_1,\ PERS{:}b_1,\ NUM{:}c_1,\ldots,\ AGR{:}o,\ldots\}$$
$$\{GEND{:}a_2,\ PERS{:}b_2,\ NUM{:}c_2,\ldots\}$$

But GKPS (36) prohibit such an analysis, by stipulating that a category-valued feature F cannot contain a category in which F itself appears. They permit only two layers of specifications.

5.4 Categories as Members of Categories

GKPS (107:fn 20), citing Basque and Makua, admit that their scheme does not cover multiple agreement of verbs, in particular DO agreement in addition to SU agreement. The problem is that V and VP each have but one member AGR, the value of which must be identical for V and VP, by the HFC--but the value of AGR for VP is 'picked up' from SU, the value of AGR for V from DO, and these will usually not be identical.

As it happens, at least two (similar, but not identical) suggestions have been made in the literature for averting this difficulty, by Anderson (1977, 1982, and elsewhere) and by Stump 1980. Both proposals view categories themselves--which is to say, sets--as potential members of categories. The result is that categories can be layered to some degree, as in the previous subsection.

In Anderson's (1977:21f) approach there is no limit to the number of layers: 'When a rule of grammar assigns features to [a morphosyntactic representation], and that [representation] already carries specifications for those features, then (unless, of course, the rule is explicitly stated so as to CHANGE the features involved, rather than simply to add them), the result is not that the new features and the old merge within the same complex, but rather that a new layer of structure is created, taking the old feature complex as its base.' In the examples Anderson considers, there is no more than one categorial member for any given category, though no assumption he makes would require this. For instance, a verb exhibiting agreement with PERS:b_1 and NUM:c_1 of SU, PERS:b_2 and NUM:c_2 of DO, and PERS:b_3 and NUM:c_3 of IO would have internal organization {N:-, V:+, BAR:0, PERS:b_1, NUM:c_1, {PERS:b_2, NUM:c_2, {PERS:b_3, NUM:c_3}}}. I'll refer to this approach as DEEP LAYERING of categories.

Stump (1980:8), on the other hand, freely permits any number of categorial members within a category, via 'the use of complex symbols of the second order. With this device, the verb agreement transformation...copies not single features, but feature complexes, which it embeds in the verb's complex symbol...Once this embedding has taken place, these feature complexes assume the role of regular features in the host complex--they are unordered, etc.' On this view, a verb exhibiting agreement with PERS:b_1 and NUM:c_1 of SU, PERS:b_2 and NUM:c_2 of DO, and PERS:b_3 and NUM:c_3 of IO would have the internal organization {N:-, V:+, BAR:0, {PERS:b_1, NUM:c_1}, {PERS:b_2, NUM:c_2}, {PERS:b_3, NUM:c_3}}. I'll refer to this approach as BROAD LAYERING of categories.

Deep layering permits different arguments of a verb to be distinguished according to the layers they occupy. It is suited to situations where more than two arguments are marked in a single form (section 4.3) and can accommodate the description of inverse forms (section 4.6), via morphological rules altering the layer assignments of specifications. Broad layering treats different arguments of a verb on a par; they are distinguishable only by virtue of the specifications within the layers (which might, as in Stump's treatment of French, include case specifications). It is suited to situations where specifications are combined (section 4.1), especially from a set of more than two arguments (section 4.2), and to the description of second-order agreement (section 4.5).

The assumptions of GKPS allow neither deep layering nor broad layering. For GKPS all members of categories are feature specifications; categories are partial functions from features to values. And it is not at all easy to see how their formalism can be amended to accommodate categories as members of categories (while preserving its other attractive qualities).

5.5 Stack-valued Features

GKPS themselves (107) declare, 'The obvious way of extending our apparatus would be to allow AGR to take a sequence of categories as value.' On this proposal, the value of AGR is a STACK, an ordered set, of categories. A verb exhibiting agreement with PERS:b_1 and NUM:c_1 of SU, PERS:b_2 and NUM:c_2 of DO, and PERS:b_3 and NUM:c_3 of IO might then have the internal organization {N:-, V:+, BAR:0, AGR:({PERS:b_1, NUM:c_1}, {PERS:b_2, NUM:c_2}, {PERS:b_3, NUM:c_3})}.

Stack-valued features would provide a way of getting the effect of deep layering within the general GKPS framework. Second-order agreement still presents a problem, however, so long as it is stipulated that AGR cannot occur 'within' AGR.

5.6 A Multiset Proposal

The proposals in 5.3-5 don't directly represent certain pieces of information about the feature specifications involved in agreement. Which specifications are inherent and which imposed by agreement rules is represented indirectly, by their position within layerings (top layer versus lower layer). In deep-layering approaches, which specifications are imposed by agreement with SU and which by agreement with DO is also represented indirectly, again by their position within layerings (first layer down versus second).

Now the motives for pursuing such frameworks are admirable. We don't want to set up a formalism that goes substantially beyond what is required for describing the facts of language--the framework should be empirically restrictive--and in any case the formalism must be interpretable and should be mathematically tractable. But I believe that the facts of agreement alone are enough to suggest that a much more general framework should be explored, and that when government is also considered (as in section 6 below) this line of theory-construction will seem even more attractive.

My suggestions here will be pre-formal, in the sense that they do not set out an actual formal system or specify its interpretation. Instead, I will sketch the outlines that I think such a formal system should have.

I propose, first, to give up the treatment of AGR as a category-valued feature parallel to SLASH, in favor of a treatment of AGR as a subset, as in sections 5.1 and 5.2 above.[5]

Next, to avert the difficulties of the set-theoretic scheme in 5.2, I will take categories not to be ordinary sets but instead to be (finite) MULTISETS,[6] that is, sets which may have repeated members. The multiset [A, A, B] is distinct from the multisets [A, B] and [A, B, B], even though the SETS {A, A, B}, {A, B}, and {A, B, B} are identical. On this view, a verb having the inherent specification NUM:c_0 and the specifications NUM:c_1 imposed by agreement with SU, NUM:c_2 by agreement with DO, and NUM:c_3 by agreement with IO would have the internal organization [N:-, V:+, BAR:0, NUM:c_0, NUM:c_1, NUM:c_2, NUM:c_3].

What, then, says that NUM:c_0 is the inherent specification here, or that NUM:c_2 is the specification imposed by DO rather than by SU or IO? After all, the order of listing is just as irrelevant in multisets as in ordinary sets; [A, B] is the same multiset as [B, A]. I will say that NUM:c_0 has the property INH and that NUM:c_2 has the properties IMP, AGR, and DO, which is equivalent to saying that NUM:c_0 belongs to the (sub)set INH, while NUM:c_2 belongs to the (sub)set IMP (and also to the subsets AGR and DO of IMP).

The intended interpretation here is that INH comprises the inherent feature specifications in a category, IMP the imposed ones; that AGR comprises the feature specifications imposed by agreement (whereas GOV comprises those imposed by government); and that DO comprises the feature specifications attributable to direct objects, either by agreement of verb with direct object or by government of direct object by verb (whereas SU, IO, and so on comprise feature specifications attributable to other grammatical relations).

The way in which individual specifications are allotted to the sets INH, IMP, AGR, and so on must itself be described. This allotment should follow, as much as possible, from universal principles plus the details of independently required rules in particular languages. Thus, morphosyntactic feature specifications belong to the set INH by universal default; only when the CAP or a government rule requires the presence of a specification is it assigned to the set IMP. Specifications required on a controlled sister category by the CAP (with reference to a controller sister category) belong to the set AGR and to the set *G1*, where *G1* is the grammatical relation that the controller bears to the controllee; while specifications required on a category by virtue of a government rule belong to the set GOV and to the set *G2*, where *G2* is the grammatical relation that the governed category bears to the governor. The HFC must preserve membership in the sets INH, IMP, etc., and I will assume that the CAP does so as well, as long as its own requirements are satisfied.

Such a scheme is adequate for most of the phenomena I have been discussing, though this might not be obvious for second-order agreement. As in section 5.3 above, consider a PD SU, of the Turkish type, in a language with subject-verb agreement. The PR complement will be a category with the specifications NUM:c_1 and NUM:c_2 in it--NUM:c_1 belonging to INH, NUM:c_2 belonging to IMP, AGR, and CP. The overall NP will have the same specifications, belonging to the same sets.

A VP sister of this NP might have its own inherent number, NUM:c_3. But in any case, the CAP imposes some specifications on this VP. In many languages, only specifications belonging to INH are imposed on verbs, but I will suppose that my hypothetical language is like Maithilī and also imposes specifications belonging to IMP. The VP will then have NUM:c_1, belonging to IMP, AGR, and SU, among its members, and it will also have NUM:c_2, belonging to IMP, AGR, SU, and CP, among its members (membership in the first three following directly from the CAP, membership in CP following from the assumption that the CAP preserves membership as far as possible). And the head V will share all of these specifications and their class memberships.

The multiset proposal can then make a variety of feature specifications, along with certain (limited) information about the 'source' of these specifications, directly available to morphological rules.

6. Imposed Versus Inherent

I now return to presenting data that bear on the choice of an appropriate formalism for features. In this section I discuss three phenomena that can be seen as motivating a formalism in which INH and IMP are explicitly represented: the resolution of syntactic feature conflict (6.1), Upper Sorbian possessive agreement (6.2), and agreement and government in Russian numeral expressions (6.3).

6.1 Resolving Syntactic Feature Conflict

Pullum and Zwicky (to appear) observe that syntactic feature conflicts involving coordination can arise in several ways. An example like **At present the project managers, but in the past the executive directors, set the research priorities* is bad because the verb *set* must be understood both as a present tense form and as a past tense form. Agreement and government can also give rise to such conflicts. A verb might have to agree in gender with a coordinate NP composed of NPs of different genders, and the case of an object NP might be governed by a coordinate V composed of Vs governing different cases.

In some instances there are general principles, referring either to semantic or to syntactic properties of the relevant material, that select a form that 'resolves' the conflict. A verb might, for example, have the specification GEND:MASC if one or more of its coordinate subjects has this specification. And resolution is always possible when a language provides inflectional forms that are neutral with respect to the difference in values--as English does with respect to the difference between PERS:1 and PERS:2 in all verbs except *be*, so that *Either you or I have the solution* presents no problems in agreement, though *Either you or I am/are responsible* does for many speakers.

As for forms that are ambiguous with respect to a difference in values, they often cannot resolve a conflict. Though a few English verbs happen to have their past tense forms identical to their present tense forms, as *set* does, this is essentially an accident, and the existence of an ambiguous form like *set* that can be understood in either way will not resolve a conflict, as I illustrated above. However, sometimes ambiguous forms WILL resolve conflicts, as the form *come* (ambiguous between infinitive and past participle) does in *I probably will, and you already have, come to understand why there is a problem here.*

Drawing on (a limited set of) data from German and Xhosa as well as English, Pullum and Zwicky argue that ambiguous forms can resolve conflicts only when the feature specifications are imposed (as in my *come* example) rather than inherent (as in my *set* example). 'Imposed' here means imposed either by a government rule, as in my *come* example

(where the form of the verb is governed by an auxiliary), or by agreement, as in a Xhosa example (where the gender class of a predicate adjective agrees with the subject) originally due to Voeltz (1971); that is, imposed specifications are those belonging to IMP as in section 5.6.

6.2 Upper Sorbian Possessives

Corbett (to appear) points out some facts about the West Slavic language Upper Sorbian that can be interpreted as involving the difference between inherent and imposed features.

As in section 2 above the issue is possessives, in this case nonpronominal possessives: possessive adjectives (PAs) like *bratrow-* 'brother's', formed from *bratr* 'brother'. Such PAs, which are understood as definite singulars, show agreement in gender, number, and case with a sister N^1. But a possessive pronoun that modifies such a PA doesn't agree with this N^1 but has its own inherent CASE:GEN and otherwise agrees with the base noun of the PA; thus the pronoun has CASE:GEN and NUM:SG and whatever value of GEND the base noun has: *mojeho* (GEN SG MASC) *bratrowe* (NOM PL) *dźěći* (NOM PL) 'my brother's children' (Corbett's (1)), *mojeho* (GEN SG MASC) *mužowa* (NOM SG FEM) *sotra* (NOM SG FEM) 'my husband's sister' (Corbett's (15)).

Apparently a way must be found to say that the possessive adjective *mužowa* has the specification CASE:GEN NUM:SG GEND:MASC in INH, but CASE:NOM NUM:SG GEND:FEM by agreement, that is, in IMP, AGR, and HD.[7] The specification CASE:GEN in INH is realized by a rule of inflectional morphology (see Zwicky 1985b on realization rules) as the suffix *-ow*, but it is the specifications in IMP that are realized in the outer inflectional suffixes on the adjective. However, only the specifications for NUM and GEND that belong to INH are imposed by agreement on a modifying pronoun.

For this analysis to be available, at least three things must be true. First, CASE:GEN must be treated as a specification in INH for certain possessives. Second, realization rules must be able to refer to specifications stipulated as being in either INH or IMP, though presumably the unmarked situation is for such rules to refer to specifications in IMP. And third, the parochial aspects of agreement (that is, the aspects of agreement not given by the CAP) must be able to include at least (a) a stipulation of which controller-controllee pairs actually participate in agreement; (b) for some individual pairs, a stipulation of which features participate in agreement; (c) for some features in some pairs, a stipulation of what happens when there is a CONFLICT between the controller's specifications in INH and in IMP--stipulating, in Upper Sorbian, that when there is a conflict between INH and IMP values of CASE, GEND, and NUM on pronominal PR in PD-PR agreement, only the INH values are imposed on PR.

6.3 Agreement and Government in Standard Russian Numeral Constructions

The cardinal numeral words (CNWs) of Russian fall into three groups with respect to the way their forms are related to the forms of the N^1 constituents with which they combine within an NP.[8]

In what I will call groups A the CNW shows the case and number that the NP bears, and it governs a fixed case and number on N^1. In group C the CNW governs a fixed number on N^1 but both CNW and N^1 show the case borne by the NP; that is, in group C the CNW agrees with N^1 in case. In an intermediate group B the CNW behaves like a group A word for certain NP cases but like a group C word in others; there are two subtypes of group B, depending on the case governed by CNW.

The assignment of a CNW to a group follows from its morphological composition, in particular from which CNW it ends with. Group A comprises the words for 1,000, 1,000,000, and 1,000,000,000 and all CNWs ending in one of these words. Group C comprises the word for 1 and all CNWs ending in this word. Group B comprises all the remaining numerals; in group B1 are the words for 2 through 4 and all CNWs ending in one of them, and in group B2 are the words for 5 through 19, the words for the tens from 20 through 90, the words for the hundreds from 100 through 900, and all CNWs ending in one of them. The interesting analytic question of how individual CNWs are assigned to groups, in particular of how to express the observation that the last word within a CNW serves as its 'head' in the sense of morphological determinant (Zwicky 1985a), is not my concern here, however.

The relationships between CNW and N^1 for the features GEND, CASE, and NUM are summarized in the table below.

Group	Exemplar	Gender	Number of CNW	Case	Number of N^1
A	1000	----	SG	governs GEN	PL
C	21	agrees	SG	agrees	SG
B	22 (B1) 12 (B2)	agrees	PL	direct: governs GEN	SG (B1), PL (B2)
				oblique: agrees	PL

6.3.1 Head-complement Constructions

The group A pattern, as in *tysjač-a* (NOM SG FEM) *dom-ov* (GEN PL MASC) 'a thousand houses', is straightforward. The CNW here is simply an N, the head N of NP, with an N^1 complement, and it governs CASE:GEN

and NUM:PL on its complement, just as Ns like *djužin* 'dozen' do. The head N and the overall NP then share their specifications for CASE, GEND, and NUM, all belonging to INH; the complement has its own GEND specification, in INH, and specifications for GEND and NUM in IMP, GOV, and CP.[9]

6.3.2 Modifier-head Constructions

The group C pattern, as in *odn-a* (NOM SG FEM) *komnat-a* (NOM SG FEM) 'one room' and *odn-im* (INST SG MASC) *dom-om* (INST SG MASC) '(with) one house', is clearly one of modifier plus head. The modifier CNW agrees in GEND and CASE with its head N^1, in the same way that adjectives do.

What is novel for group C CNWs, as against adjectives, is that they idiosyncratically (and counter-semantically) both HAVE the specification NUM:SG and GOVERN it on their heads: *dvadcat' odin dom* (NOM SG MASC) 'twenty-one houses'. This specification might be analyzed as a consequence of government of NUM:SG on heads, in combination with agreement of modifiers with their heads with respect to NUM as well as CASE and GEND. If so, then we must stipulate how a conflict between specifications in INH and in IMP is resolved, since the head will have both the specification NUM:PL in INH (shared with the NP dominating it; specifications on NUM on NPs belong in general to INH) and also the specification NUM:SG in IMP (by virtue of number government). In this instance the conflict is resolved in favor of the specification in IMP, with the result that the CNW has the specification NUM:SG in IMP (and also in GOV, AGR, and HD).

As a result, group C CNWs would be treated as doubly special, in governing NUM:SG and in agreeing with IMP specifications if there is a conflict.

I now turn to the B pattern, illustrated below. This is the pattern that has excited the interest of analysts; see in particular Babby (1984, 1985) and the works cited there.

B1		B2		
dv-a	dom-a	pjat'	dom-ov	
NOM PL MASC	GEN SG MASC	NOM PL	GEN PL MASC	
'two houses'		'five houses'		
				DIRECT
dv-e	komnat-y	pjat'	komnat	
NOM PL FEM	GEN SG FEM	NOM PL	GEN PL FEM	
'two rooms'		'five rooms'		

B1		B2		
dv-um	dom-am	pjat-i	dom-am	
DAT PL	DAT PL MASC	DAT PL	DAT PL MASC	OBLIQUE
'(to) two houses'		'(to) five houses'		

First, gender. The only group B CNW with distinct gender forms is 2: *dva* (MASC/NEUT), *dve* (FEM). As the phrases above indicate, group B CNWs act like modifiers in showing agreement with respect to this feature. I will in fact assume that the group B pattern is modifier-head, as in group C, and no head-complement as in group A.

Number and case in the oblique cases (GEN, DAT[ive], INST[rumental], and PREP[ositional], versus NOM and ACC) are HOMOGENEOUS, to use Babby's terminology. The CNW and its head share these features. In the direct cases, number and case are HETEROGENEOUS. Group B CNWs are apparently inherently PL (though the B2s are of a rather eccentric declensional type). In any event, the B1 CNWs clearly govern NUM on their heads, since the SG value of this feature on their heads is neither agreeing nor semantically appropriate. Nothing is lost by assuming that the B2s also govern NUM on their heads. For either subgroup, then, the head has a specification for NUM in IMP, GOV, and HD.

But now there are conflicts with respect to NUM and CASE. First, the value of NUM in CNW: The head N^1 has one value for NUM in INH (PL, shared with its dominating NP) and another in IMP, GOV, and HD (determined by its CNW governor--SG for B1s, PL for B2s), and when there is a conflict the CNW agrees with the specification in INH, NUM:PL.

As for the conflict with respect to NUM in the head itself, here resolution is bound up with case assignment. I will assume, with Babby, that the direct cases NOM and ACC are 'structural' cases--in my terms, that their specifications for CASE belong to INH rather than IMP--whereas the oblique cases are assigned by government rules, and so have specifications for CASE belonging to IMP, GOV, and *G*, where *G* is a grammatical relation other than HD. It then follows that in the direct cases the head N^1 and its head N have only one specification for CASE in IMP and GOV (a specification in HD, imposed by the CNW), while in the oblique cases the head N^1 and its head N have TWO specifications for CASE in IMP and GOV--an INTERNAL one in HD, imposed by the CNW, and an EXTERNAL one (in a non-HD set) shared with the dominating NP.

Resolution of CASE conflicts is then a matter of conditions on realization rules: realization of an external CASE specification overrides realization of an internal one (this predicts that external CASE is homogeneous); realization of a CASE specification in IMP overrides realization of one in INH; otherwise a CASE specification in INH is realized. This hierarchy for realization applies both to Ns and to

CNWs. Resolution of NUM conflicts is also a matter of realization rules: in the presence of an external specification of CASE, realization of a NUM specification in INH overrides realization of one in IMP (this predicts homogeneous NUM to go along with homogeneous CASE); otherwise realization of a NUM specification in IMP overrides realization of one in INH; otherwise a NUM specification in INH is realized.

I now summarize what might appear to be an extraordinarily complex analysis for the modifier-head configurations involving CNWs.

I. All modifier CNWs agree in NUM, CASE, and GEND with their heads; if there is a conflict between specifications in IMP and INH, group C CNWs agree with NUM in IMP and group B CNWs with NUM in INH.

II. All modifier CNWs govern the NUM of their heads; group C and B1 CNWs govern NUM:SG, and group B2 NUM:PL.

III. Group B CNWs govern CASE:GEN on their heads.

IV. Realization of external CASE specifications overrides realization of internal specifications; and in the presence of external CASE, realization of NUM in INH overrides realization on NUM in IMP.

V. Realization of specifications in IMP overrides realization of corresponding specifications in INH, the latter serving as defaults.

If it were not for the fact that modifier CNWs govern NUM and CASE on their heads (principles II and III), the system would be childishly simple, involving agreement and government of the most ordinary sort, with conflicts resolved by the general principle V. However, the existence of NP-internal government in addition to NP-external government gives rise to a number of further conflicts in specifications, and these are resolved by the rider on principle I and by the principles in IV.[10]

7. Final Observations

I hope to have laid the groundwork for a multiset approach to feature specifications in generative grammar and to have motivated some of its details, in particular a formal distinction between IMP[osed] and INH[erent] specifications, and within IMP, between AGR[eement] and GOV[ernment] specifications. This approach also incorporates a limited formalism for reference to grammatical relations like SU, DO, IO, HD, and CP, though many details of this part of the formalism remain to be worked out. Finally, in this framework the analysis of inflectional

morphology is divided between principles of syntax--in particular, the CAP, along with parochial conditions as to when this principle applies, and rules of government-- and principles of morphology, including realization rules sensitive to properties like INH, IMP, GOV, AGR, CP, and HD.

The data surveyed here seem to me to give no encouragement to theorists who parsimoniously propose to collapse two, or all three, of the notions SUBCATEGORIZATION, GOVERNMENT, and AGREEMENT, and little encouragement to standard GPSG positions on the nature of AGR. Instead, I suggest that formal elegance might in this instance have to be sacrificed to empirical adequacy.

NOTES

*This paper was written at Stanford in winter quarter 1986. My thanks to Steve Anderson, Jonni Kanerva, and Geoff Pullum, who provided substantive and bibliographic help without realizing what use I was going to make of it.

[1]Both patterns of agreement are widespread; Moravcsik (1971:A61f) lists (among others) various Indo-European languages, Coptic, Fijian, Fulani, and Hausa in the first set, and various Algonquian and Finno-Ugric languages, Chinook, Eskimo, Hopi, and Tzeltal in the second.

[2]See Lewis (1967:38-43), Underhill (1976:90-3) for details.

[3]I have argued elsewhere, in Zwicky (to appear), that agreement and subcategorization should not be collapsed, because the sort of 'phonological resolution' studied by Pullum and Zwicky (to appear), and discussed in section 6.1, does not apply to subcategorization conflicts, only to agreement and government conflicts.

[4]The connection between the two types of agreement is explicit in Keenan (1974:sec. 4.1.2).

[5]Whether SLASH itself is best treated as a category-valued feature is a separate issue.

[6]See GKPS (531) for a use of multisets in a different context.

[7]Sadock (1985:sec. 5) takes a very different view of the situation in Upper Sorbian, proposing that a phrase like *mojeho bratrowe* has a morphological analysis, into *mojeho bratr* plus the affix *-owe*, that diverges from its syntactic analysis.

[8]The sketch of the Russian facts that follows is based on Maltzoff 1984.

[9]I am not proposing to argue here for a particular list of grammatical relations, but I must at least assign labels to the ones involved in my examples.

[10]I have intentionally left a number of further complications--among them, modifying adjectives and prepositional quantifiers--out of this picture, which was never intended to be a full analysis of Russian agreement and government. See the studies by Babby for discussion of these complications and their theoretical significance.

REFERENCES

Anderson, Stephen R. 1977. On the formal description of inflection. *CLS* 13.15-44.

_____. 1982. Where's morphology? *Linguistic Inquiry* 13.571-612.

_____. 1985. Inflectional morphology. In Timothy Shopen (ed.), *Language Typology and Syntactic Description. Vol. III: Grammatical Categories and the Lexicon.* Cambridge University Press, pp. 150-201.

Babby, Leonard H. 1984. Case conflicts and their resolution: A contribution to EST case theory. *Cornell Working Papers in Linguistics* 6.1-21.

_____. 1985. Noun phrase internal case agreement in Russian. *Russian Linguistics* 9.1-15.

Corbett, Greville. To appear. The morphology-syntax interface: Evidence from possessive adjectives in Slavonic. *Language* xx.xxx-xxx.

Durie, Mark. 1986. The grammaticization of number as a verbal category. To appear in *BLS* 12.

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

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

Keenan, Edward L. 1974. The functional principle: Generalizing the notion of 'subject of'. *CLS* 10.298-309.

Jhā, Subhadra. 1958. *The Formation of the Maithilī Language.* London: Luzac & Co.

Lewis, G. L. 1967. *Turkish Grammar.* Oxford: Oxford University Press.

Maltzoff, Nicholas. 1984. *Essentials of Russian Grammar.* Lincolnwood, IL: National Textbook Co.

Moravcsik, Edith A. 1971. Agreement. *Stanford WPLU* 5.A1-A69.

Pullum, Geoffrey K. and Arnold M. Zwicky. To appear. Phonological resolution of syntactic feature conflict. *Language* xx.xxx-xxx. (Earlier version: Syntax Research Center, University of California at Santa Cruz, SRC-85-03 (1985)).

Sadock, Jerrold M. 1985. Autolexical syntax: A theory of noun incorporation and similar phenomena. *NLLT* 3:4.379-439.

Stump, Gregory T. 1980. An inflectional approach to French clitics. *Ohio State University Working Papers in Linguistics* 24.1-54.

Underhill, Robert. 1976. *Turkish Grammar*. Cambridge MA: MIT Press.

Voeltz, Erhard. 1971. Surface constraints and agreement resolution: Some evidence from Xhosa. *Studies in African Linguistics* 2.37-60.

Zwicky, Arnold M. 1977. Hierarchies of person. *CLS* 13.714-33.

_____. 1985a. Heads. *Journal of Linguistics* 21:1.1-30.

_____. 1985b. How to describe inflection. *BLS* 11.372-86.

_____. To appear. German adjective agreement in GPSG. *Linguistics* xx.xxx-xxx.