

CONSTRAINTS ON AGREEMENT[1]

Greville G. Corbett

University of Surrey, UK

Abstract: Agreement shows massive variation (across languages and within languages), caused largely by the conflicting effects of syntax and semantics. Copying accounts of agreement have been shown to be inadequate. Unification-based accounts, where agreement can be seen as cumulating partial information from the controller and the target, have much better prospects. However, the variation induced by semantic/syntactic clashes is subject to well-established constraints, such as the Agreement Hierarchy and the Predicate Hierarchy, but these constraints do not fit readily into current accounts of agreement.

Keywords: agreement, constraints, hierarchies, unification, variation

[1] The support of the British Academy and of the ESRC (grant R000236063) is gratefully acknowledged. I am grateful to Ivan Sag for useful comments on some of the issues raised, and to Dunstan Brown and Norman Fraser for helpful discussion of a draft. A version of this paper was read at the panel "Linguistic Theory and Language Universals", organized by Bernard Comrie and Maria Polinsky, under the auspices of the 16th International Congress of Linguists, Paris, July 1997. I would like to thank those who attended the session for their suggestions.

ISBN: 0 08 043 438X

0. INTRODUCTION

Agreement is a central concern for those interested in constraints, since it shows massive variation (across languages and within languages), it is of considerable current concern in linguistic theory, the variation which we observe shows universal regularities BUT current theories seem unable to capture these regularities. I shall show how the nature of agreement provides arguments for a constraint-based approach to syntax, but that data which have been around for a considerable time still prove problematic even for constraint-based approaches. I will consider the nature of agreement (section 1) and point out briefly the advantages of unification-based approaches for its analysis (section 2). I will then present data on agreement mismatches, where we find variation but very regular variation (section 3). This proves seriously problematic for current approaches. Finally I suggest possible ways forward (section 4).

1. THE NATURE OF AGREEMENT

Agreement is of course reflected in morphology, and there are unexpected complications there (Corbett forthcoming). But the nub of the constraints problem concerns rather what should determine the agreement features realized by the morphology. The cause of much of the variation we shall observe is that agreement is a matter both of syntax and of semantics.

Evidence that semantics is involved comes from (British) English plural agreement with nouns like *committee*. Such nouns are singular morphologically, and they have a normal plural. And yet, when singular, they may take plural agreement:

(1) The committee have decided.

This possibility depends on the semantics of the noun. Yet there is much more to be said, because the likelihood of examples like (1) being found depends on the variety of English involved: British, American, Australian and New Zealand (Nixon 1972, Johansson 1979, Watson 1979, Bauer 1988). Consider the following judgements, when subjects were asked to consider sentences apparently produced by non-native speakers of English and to correct them where necessary (data from Johansson 1979: 203, Bauer 1988: 254). The test sentence which concerns us is:

(2) The audience were enjoying every minute of the show.

Table 1: Agreement with 'corporate' nouns in three varieties of English

		variety		
		GB (N=92)	US (N=93)	NZ (N=102)
response (%)	no correction	77.2	5.4	72.5
	was enjoying	15.2	90.3	20.6
	other response	7.6	4.3	6.9

There is substantial divergence between the varieties in the choice of agreement. Another source of variation is the agreement target; while Table 1 is concerned with the predicate, it is known that the relative frequency of plural agreement is different for other agreement targets such as the personal pronoun, as we shall see in section 3.1. From this it will be clear that I am using 'agreement' in the wider sense to include the determination of the form of anaphoric pronouns as a part of agreement. Most mainstream work on agreement uses the term in this sense, to include pronouns. Barlow (1988:134-52; 1991) reviews the literature and concludes that there are no good grounds for distinguishing between agreement and antecedent-anaphora relations.

ISBN: 0 08 043 438X

Evidence such as that presented so far has convinced many linguists that there is indeed a place for semantics in determining agreement. There is a temptation, which linguists yield to from time to time, to treat agreement as entirely a matter of semantics. This view appears to me to be untenable. Consider first Herb Clarke's example (cited in Barlow 1988: 227)

(3) I'm parked on the hill / *I is parked on the hill

The first form is acceptable for *my car is parked on the hill,* but semantic agreement is impossible here: **I is parked on the hill.*[2] A further instance where a form which might be expected on semantic grounds is excluded was pointed out by Perlmutter (1972: 245)

(4) *The hedgehog are becoming extinct.

A different type of evidence concerns sentences where two different agreements are found together:

(5) This man and woman were on their way to the shops, when ...

Semantics would not explain both possibilities in (5), that is, singular agreement within the noun phrase but plural agreement of the predicate. Note that we cannot simply exclude the noun phrase from the domain of semantic control (which would in any case be an odd restriction) since there are cases similar to (5) where plural agreement (as expected from semantic considerations) is possible:[3]

Russian

(6)

Marija	zadumalas´	ob	ostavlenn-yx	muže	i	dočeri:
Maria	thought	about	left.behind-PL	husband	and	daughter

kak	oni	tam,	čto	s	nimi?
how	they	there	what	with	them?

'Maria thought about the husband and daughter she had left behind, and wondered how they were and what was happening to them.' (Maksimov, *Karantin)*

Expressions consisting of conjoined singular noun phrases in Russian take singular agreement within the noun phrase more often than plural, but examples like (6), with plural agreement, are fully acceptable.

A further difficulty (probably insurmountable) for semantics-only approaches to agreement is provided by the South Caucasian language Georgian (Aronson 1982: 243, 406-407, quoted in Durie 1986):

Georgian

(7)

ivane	še-mo-vid-a	da	da-ǰd-a
John	PRV-PRV-enter-AOR.3.SG	and	PRV-sit.SG-AOR.3.SG

'John entered and sat down' (PRV = preverb, AOR = aorist)

(8)

čem-i	mšobl-eb-i	še-mo-vid-nen	da	da-sxd-nen
my-AG	parent-PL-NOM	PRV-PRV-enter-AOR.3.PL	and	PRV-sit.PL-AOR.3.PL

'My parents entered and sat down'
(AG indicates an agreement marker; the ending *-i* is syncretic, covering nominative singular and plural, and genitive singular and plural)

---

[2] But see further Pollard and Sag (1994: 85n21).

[3] Here and elsewhere glossing is provided to help make the point at issue clearer; it is not full glossing.

ISBN: 0 08 043 438X

The verbs agree in number in a straightforward way. Additionally, though, the verb 'sit' has different forms according to whether one person sits (*daǰd-*), or more than one *(dasxd-),* unlike the verb 'enter'. The choice of verb forms can be seen as a case of verbal number, determined by semantic considerations. Now consider what happens when there is a numeral phrase. Numerals require a singular noun (*megobari*, the plural would be *megobr-eb-i)* and the resulting phrase controls singular agreement (already somewhat problematic for a semantic account):

(9)	čem-i	sam-i	megobar-i	še-mo-vid-a	da
	my-AG	three-AG	friend.SG-NOM	PRV-PRV-enter-AOR.3.SG	and

da-sxd-a
PRV-sit.PL-AOR.3.SG
'My three friends entered and sat down'

Singular agreement is found on both verbs. Yet the second, which has two forms according to the number of participants, shows the plural verbal form *dasxd-*, since more than one participant is involved in the action. In other words, the verb's form is determined as plural in terms of verbal number, by the semantics, but this does not determine the agreement, which is singular. (For further arguments against a 'semantics only' view of agreement see Pollard and Sag 1994: 71-73, 98.)

I conclude that agreement cannot be handled just in syntax, nor just in semantics, and that the traditional view giving a role to both is correct.

2. THE ADVANTAGE OF UNIFICATION

The familiar rule-based approaches to agreement copied features from controller to target. These feature-copying approaches face several problems: the controller may be absent (as in pro-drop languages), or it may be present but be underspecified, something which occurs frequently with pronouns (Barlow 1988: 30-43; his arguments are developed in Pollard and Sag 1994: 62-67). Unification-based accounts, where agreement can be seen as a matter of cumulating partial information from the controller and the target, have much better prospects (Shieber 1986: 21-22, Barlow 1988: 22-45, but see Bayer and Johnson 1995 for problems). Consider just one of the arguments, starting from the following example:

French

(10)	Je	suis	content/content-e
	I	be.1ST.SG	pleased.SG.MASC/pleased-SG.FEM

'I am pleased' (man/woman talking)

In accounts based on a rule of feature-copying, we need to say that French has two pronouns *je,* one masculine and one feminine, which happen to be phonologically identical. In a unification-based approach, we could have the following feature structures (the first for the pronoun and the verb, and the second for the predicative adjective):

(11)
$$\begin{bmatrix}\text{number: singular}\\ \text{person: 1st}\end{bmatrix} \qquad \begin{bmatrix}\text{number: singular}\\ \text{gender: feminine}\end{bmatrix}$$

These feature structures can be unified, since they are compatible, to give the following structure:

ISBN: 0 08 043 438X

(12)
$$\begin{bmatrix} \text{number: singular} \\ \text{person: 1st} \\ \text{gender: feminine} \end{bmatrix}$$

Thus the information is cumulated from different parts of the structure. These approaches to language may be called 'constraint-based approaches' (Shieber 1992: 1); they specify, as constraints, that particular feature structures must unify.

However, specifically for representing agreement, this leaves the question of asymmetry. There is a strong intuition, captured in the controller - target terminology, that agreement is asymmetrical. In Generalized Phrase Structure Grammar this notion is reintroduced by the Control Agreement Principle (based on Keenan 1974), which specifies possible controllers and targets, and gives them different statuses (see Gazdar, Klein, Pullum and Sag 1985). In Head-Driven Phrase Structure Grammar the asymmetry is captured through 'anchoring'; gender, number and person features are anchored to real world entities through noun phrase indices, even though they may be expressed morphologically other than on the noun phrase (see Pollard and Sag 1994: 60-99, and compare Kathol forthcoming).

3. REGULAR VARIATION

The recurrent problem is that we find mismatches - instances where controller and target may realize feature values which do not unify neatly as in the French example (10), but which are contradictory, as with English *'committee*-type' nouns (such as *audience* in example 2)). Possible types of controllers which may be involved are given in Table 2:

Table 2: Types of controllers which induce agreement mismatches

controller type	example
unique or virtually unique lexical item	Serbo-Croat *deca* 'children'
set of semantically similar lexical items	English *committee*-type nouns
lexically restricted construction	masculine nouns quantified by numeral '2', '3' or '4' in Serbo-Croat
construction	conjoined noun phrases (many examples)

3.1 The Agreement Hierarchy

The agreement mismatches we find are extremely varied at first sight, yet they show quite clear patterns. Consider, for example, our third category in Table 2. We are concerned with masculine nouns in Serbo-Croat, when quantified with the numerals '2', '3' and '4'. The noun itself stands in a special form, a survival of the dual number which is synchronically a genitive singular. Attributive modifiers must take the ending *-a*. This is a remnant of the dual number; there are arguments for analysing it synchronically as a neuter plural (Corbett 1983a: 13-14, 89-92); here I will just label it as 'remnant'.

Serbo-Croat

(13) dva dobr-a čovek-a
two good-REMNANT man-SG.GEN
'two good men'

In the predicate the remnant form is again found, but so is the masculine plural form, the one we might have expected:

ISBN: 0 08 043 438X

(14) ova dva čovek-a su dobr-a/dobr-i
these-REMNANT two man-SG.GEN are good-REMNANT/good-PL.MASC
'these two men are good'

The relative pronoun is also found in both forms:

(15) dva čovek-a koj-a/koj-i ...
two man-SG.GEN who-REMNANT/who-PL.MASC ...
'two men who ...'

The personal pronoun must stand in the masculine plural form *oni* (**ona* is unacceptable). We may treat the remnant form as syntactic agreement, determined by formal factors, while the masculine plural is an instance of semantic agreement (the influence of semantic factors is evident). We find syntactic agreement in attributive position, both types of agreement occur in the predicate and the relative pronoun, and only semantic agreement is found with the personal pronoun. There is statistical data on the distribution of the forms, in the two domains where there is an option, collected before the claim I wish to make was first put forward. The figures are derived from Sand (1971: 55-56, 63) and presented in Table 3:

Table 3: Percentage distribution of 'remnant' and masculine plural forms in Serbo-Croat

	attributive	predicate	relative pronoun	personal pronoun
percentage showing plural (semantic) agreement	0	18 (N=376)	62 (N=32)	100

For completeness, Table 3 also includes the positions where there is no choice. We can see that each successive cell in the table shows a monotonic increase in the likelihood of agreement forms with greater semantic justification. This is just one instance of many. It is part of the evidence for the Agreement Hierarchy. As we have just seen, four types of agreement targets can be distinguished:

attributive < predicate < relative pronoun < personal pronoun

Figure 1: The Agreement Hierarchy

The Agreement Hierarchy allows us to constrain possible agreement patterns as follows:

> For any controller that permits alternative agreement forms, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement forms with greater semantic justification will increase monotonically (that is, with no intervening decrease).

A small sample of the supporting evidence is given in Table 4 (see Corbett 1979, Corbett 1983: 8-41, 1991: 225-260 for further evidence).

ISBN: 0 08 043 438X

Table 4: The Agreement Hierarchy: distribution of gender and number agreement

	attributive	predicate	relative pronoun	personal pronoun
Chichewa diminutive for human	class 12	class 12	class 12	class 12/ (CLASS 1)
English *committee*	sg	sg/PL	sg/PL	sg/PL
Serbo-Croat 2-4 plus masculine noun	remnant	remnant/(M PL) (M PL 18%)	(remnant)/M PL (M PL 62%)	M PL
Russian conjoined NPs	sg/(PL) (PL 14%)	(sg)/PL (PL 71%)	(sg)/PL	(sg)/PL

Note: lower case indicates syntactic agreement, and upper case SEMANTIC AGREEMENT; parentheses indicate a less frequent variant.

3..2 The Predicate Hierarchy

So we see that the variation induced by semantic/syntactic clashes is subject to constraints; we see this in the Agreement Hierarchy, and if we home in on one segment of it, the predicate, we find a similar picture, as was pointed out by Comrie (1975). He showed how honorific plural pronouns may take singular or plural agreement, with considerable variation, but that this variation is constrained by what I shall call the 'Predicate Hierarchy':

verb < participle < adjective < noun

Figure 2: The Predicate Hierarchy

Reformulating Comrie's proposal we may claim that:

> For any controller that permits alternative agreement forms, as we move rightwards along the Predicate Hierarchy, the likelihood of agreement forms with greater semantic justification will increase monotonically (that is, with no intervening decrease).

In subsequent research I checked on all the Slavonic languages, for agreement with honorific pronouns, and the results (given in detail in Corbett 1983a: 42-59) are given in summary form in Table 5.

ISBN: 0 08 043 438X

Table 5: Agreement with honorific *vy* in the Slavonic Languages (Corbett 1983a: 56)

	finite verb	participle	adjective	noun
West Slavonic:				
Czech	pl	(pl)/SG	(pl)/SG	SG
Slovak	pl	pl/(SG)	SG	SG
Lower Sorbian	pl	pl	pl/SG	SG
Upper Sorbian	pl	(pl)/SG	(pl)/SG	SG
Polish dialects	pl	pl/SG	pl/SG	SG
South Slavonic:				
Bulgarian	pl	pl (96%) N=167	SG (97%) N=163	SG
Macedonian	pl	pl	(pl)/SG	SG
Serbo-Croat	pl	pl	pl/(SG)	SG
Slovene	pl	pl/(SG)	pl(SG)	SG
East Slavonic:				
Ukrainian	pl	pl/(SG)	(pl)/SG	SG
Belarusian	pl	pl	SG	SG
Russian	pl	pl	short form pl (97%) N=145 / long form SG (89%) N=37	SG

Note: lower case indicates syntactic agreement, and upper case SEMANTIC AGREEMENT. The sources of the percentage figures are given in the source quoted. Other parentheses indicate less frequent or less preferred variants.

Again there is great variation, but the overall pattern is very clear.[4]

3..3 Agreement with quantified phrases

As a final example, consider agreement with numeral phrases (more generally than in the special Serbo-Croat construction analysed earlier). Given a phrase consisting of numeral and noun, we might expect that the agreements would be plural (or, say, dual where appropriate), following the semantics. Or that they would be singular (as in (9) from Georgian), either as a result of failed agreement with a non-prototypical noun phrase or because number is already indicated by the quantifier and so further marking is unnecessary. Some languages, however, allow both possibilities. Let us consider briefly the Slavonic family in this light. The following Russian examples are both fully acceptable:

Russian

(16) vošl-o pjat´ devušek
came.in-SG.NEUT five.NOM girl.PL.GEN
'five girls came in'

(17) vošl-i pjat´ devušek
came.in-PL five.NOM girl.PL.GEN
'five girls came in'

The choice is affected by the animacy of the subject, and its position relative to the predicate. But the quantifier itself also has a substantial influence; Table 6 gives data on the different Slavonic languages.

[4] Naturally we should consider how the two hierarchies combine. This is not straightforward: the Predicate Hierarchy forms a sub-hierarchy within the Agreement Hierarchy. For data and discussion see Corbett (1983a: 76-93).

ISBN: 0 08 043 438X

Table 6: Predicate Agreement with Numeral Phrases in Slavonic

	2	3	4	5-10	100
West Slavonic:					
Czech	PL	PL	PL	sg	sg
Slovak	PL	PL	PL	PL /sg	sg
Sorbian	DUAL	PL	PL	PL /sg	sg
Polish	99% PL (N=123)	91% PL (N=43)	‖100% PL (N=15)	7% PL (N=68)	
South Slavonic:					
Old Church Slavonic	DUAL	PL	PL	(PL)/sg	
Bulgarian	PL	PL	PL	PL	PL
Macedonian	PL	PL	PL	PL	PL
Serbo-Croat	97% PL (N=735)	89% PL (N=249)	83% PL (N=133)	7% PL (N=1161)	
Slovene	DUAL	PL	PL	sg	sg
East Slavonic:					
Ukrainian	83% PL (N=208)	79% PL (N=150)	74% PL (N=34)	38% PL (N=45)	21% PL (N=14)
Belarusian	92% PL (N=219)	78% PL (N=67)	63% PL (N=16)	39% PL (N=49)	‖ 50% PL (N=2)
Russian	86% PL (N=541)	77% PL (N=247)	76% PL (N=68)	50% PL (N=220)	

NOTE: DUAL (where available) and PLURAL represent semantic agreement

When an entry in Table 6 consists of a single abbreviation (e.g. 'PL'), this indicates that the form given is used in the majority of cases, though not necessarily all. Thus in Slovene, the plural is normal with '3' and '4', but the singular may be used in expressions of time. Where there are no more precise data, these few exceptions are ignored (time expressions also account for some of the singular forms with '2', '3' and '4' in other languages). The languages are ordered roughly in order of decreasing use of the plural. A gap indicates a lack of data.[5]

[5] The judgements and statistics presented are taken from Suprun (1969: 175-187) unless otherwise stated. In Slovak, with the numerals '5-10' the plural is used with masculine personal forms and otherwise the singular; exceptions amount to less than one per cent of the examples, according to Suprun. *Sto* '100' takes the singular (Ján Bosák and Ľubomír Ďurovič personal communications). Sorbian preserves the dual number; otherwise agreements are broadly similar to those of Slovak (Suprun 1963a). Old Church Slavonic data are given in Suprun (1961: 81-86); for the '5-10' entry, he has ten examples of singular predicates, six of plural predicates and two where one source has singular agreement and another has plural. However, Večerka (1960: 197) says that in the Gospels the singular is used in the overwhelming majority of instances, hence the plural entry is bracketed. Suprun also gives an example with *sъto* '100', one with *tysęšta* '1,000' and one with *tьma* '10,000' - all three with singular agreement. The Polish figures are calculated on the basis of examples given in Suprun (1963b); instances where the numeral itself is in the genitive are excluded. The final Polish entry is for numerals of all types from '5' up to '999'. There are also seven examples of agreement with *tysiąc* '1,000' - all singular. The Serbo-Croat statistics are taken from Sand (1971: 51-2, 73); the figure for *dva* 'two' includes examples with *oba* 'both'; '2-4' include compound numerals ending in '2-4' and the remaining figure is for all other numerals

ISBN: 0 08 043 438X

Let us turn to the patterns revealed by Table 6. The South Slavonic languages Bulgarian and Macedonian differ from the others in using the plural with all numerals in almost all instances. Other Slavonic languages use both singular and plural. The remaining South Slavonic languages (Old Church Slavonic, Serbo-Croat and Slovene) use the dual (when available) with phrases with the numeral '2'; otherwise they show a strong preference for plural agreement for quantified phrases with the numerals '2-4', and for singular agreement (though with varying degrees of tolerance towards the plural) with numerals from '5' upwards. In several languages the distinction between '2-4' on the one hand, and '5' upwards on the other, is fairly sharp. However, the statistics for Serbo-Croat and Polish show that here the division is not absolute. It is in these languages, together with those of the East Slavonic group, where the situation is more fluid, that we find the most interesting data. The overall picture is clear: the higher the numeral the more likely is singular agreement. The form which is semantically justified becomes more likely the lower the numeral. This is clearly true in the straightforward cases like Slovak. The statistical data too support this claim, apart from two minor inconsistencies (indicated in the table with ||): the 100 per cent plural figure for '4' in Polish, and the 50 per cent plural for '100' in Belarusian. These two cases need not concern us because the sample size for both numerals is small. Even apart from these two instances, it is not the case that there is a statistically significant difference between every pair of successive numerals in each language. However, a statistical advice is that the pattern is so overwhelming that statistical tests of significance are superfluous. What is important is that, apart from the two exceptions mentioned, the rank order of the numerals according to the frequency with which they take plural agreement is the same in the different languages and that this order is inversely related to numerical value. There is strong evidence that the lower the numeral is, the more likely it is to take semantically justified agreement.

We have seen three areas where the pattern is clear, the data constrained by the Agreement Hierarchy, the data constrained by the Predicate Hierarchy and now with predicate agreement with numeral phrases. But these constraints, which have substantial cross-linguistic support, do not fit readily into current accounts of agreement (Barlow 1991, Pollard and Sag 1994: 58, Kathol forthcoming).[6] Let us concentrate on just one, the Agreement Hierarchy. When developing a more refined account of agreement, a common first move, as Barlow (1991 points out) is to split agreement into two different phenomena: for instance, in LFG it is grammatical vs anaphoric agreement (Bresnan and Mchombo 1987).

Whatever the merits of splitting agreement into two phenomena, this does not solve our problem. We cannot simply say that, for instance, where there is a choice of agreement options, semantically justified agreement will be found within the structures which LFG calls anaphoric agreement. First, because the divide between syntactic and semantic agreement is not necessarily clear-cut. We saw this with number agreement in Serbo-Croat (Table 4), there is a choice in predicate position and for the relative pronoun. Second, at the extremes, there can still be a choice: the noun phrase must surely come under 'grammatical agreement' if agreement is split, and yet we can find semantic agreement here (example 6); conversely the personal pronoun would be expected to fall under anaphoric agreement, and yet syntactic agreement can be found here. This can be seen in French, particularly in earlier French, which used various honorific titles. These could take feminine agreement (since the nouns in their normal use were feminine) and masculine agreement, since they were used of males. Even in the personal pronoun, feminine (syntactic) forms dominated. The following example would be normal, according to Grevisse (1964: 405-406):

---

above '4'. Judgements on Slovene are from Vincenot (1975: 196) as well as from Suprun (1969: 176). The final entry for Ukrainian includes examples with *sorok* '40' as well as *sto* '100'.

[6] See also Kirby (1996: 114-119) for recent discussion of the Agreement Hierarchy in terms of the emergence of language universals.

ISBN: 0 08 043 438X

French

(18) Votre Majesté partira quand elle voudra.
your majesty leave.FUT when she wish.FUT
'Your Majesty will leave when he (literally 'she') wishes.'
(Voltaire, quoted by Grevisse 1964: 406)

The feminine pronoun is used, even though the king is addressed. However, examples with a masculine pronoun also occur:

(19) Sa Majesté fut inquiète, et de nouveau il envoya
His.FEM Majesty was worried.FEM and of new he sent

La Varenne à son ministre.
La Varenne to his minister

'His Majesty was worried, and again he sent La Varenne to his minister.'
(J. & J. Tharaud, quoted by Grevisse 1964: 405)

Examples like (19) are less usual, those like (18) were normal. This shows that syntactic agreement is possible, and in this case preferred, even for the personal pronoun.

And finally, it is not at all clear that the problem is being tackled in terms of the right grammatical component. The constraints we have been discussing, in particular the Agreement Hierarchy, are violable at the 'sentence level'. The Agreement Hierarchy does not necessarily rule out specific sentences, as this Serbo-Croat example shows.

Serbo-Croat

(20) Dva tima, koj-a se nalaze u donjem delu tabel-e,
two team, which-REMNANT REFL find in lower part table-GEN,

Radnički i Olimpija, u Kragujevcu na teškom terenu
Radnički and Olimpija, in Kragujevac on difficult pitch

igral-i su prljavo i nesportski.
played-PL.MASC AUX.PL dirtily and unsportingly

'Two teams, which find themselves in the lower part of the (league) table, Radnički and Olimpija, on a difficult pitch in Kragujevac played dirtily and unsportingly.'
(*Politika* 9.XII.1969)

Here we have semantic agreement of the predicate, but syntactic agreement of the relative pronoun. Though this is possible in individual sentences, it would not be possible generally. That is to say, the likelihood of agreement forms with greater semantic justification must be greater with the relative pronoun than with the predicate. That is the case with this type of controller in Serbo-Croat (as the data in Table 3 demonstrate), but instances like (20) are not ruled out.

That is not to deny the existence of constraints on the cooccurrence of different agreement options. For instance, Pollard and Sag (1994: 71) claim:

> "Interestingly, in many contexts either the aggregate or the nonaggregate mode of individuation is possible. However, once the choice is made within a particular local context, the mode of individuation is immutable. Thus we find examples like (30a, b), but not (30c) or (30d):
> (30) *a.* The faculty is voting itself a raise.
> *b.* The faculty are voting themselves a raise.
> *c.* *The faculty is voting themselves a raise.
> *d.* *The faculty are voting itself a raise."

ISBN: 0 08 043 438X

Here 'local context' indicates a construction where the principles of grammar require identity of indices: coargument reflexive binding and pronoun binding by a quantifier. But if we move to 'ordinary' pronominal coreference, a construction not so different from those discussed by Pollard and Sag,[7] we find again that different agreement options can cooccur, that no absolute identity requirement will be tenable. (For Pollard and Sag, in contexts where there is no constraint forcing coindexing, there are two vehicles for conveying referential identity: coindexing and coreference of distinct indices. It is the latter possibility which allows them to handle the cases where different agreement options can cooccur.)

As examples, consider these cases found by Juul (1972: 20):

(21) The Institute propose to appoint a Staff Tutor, to be mainly concerned with its Certificate for Approved School Masters. (*New Statesman* 31.III.1961, p. 530)

(22) Esso Petroleum have a vacancy for an Economist in the Economics Division of its Treasurer's Department. (*The Sunday Times* 3.III.1970, p. 41)

(23) It is most unlikely that Courtaulds will sell off any of its interests in Snia Viscosa, ... Courtaulds feel that this would merely be pawning its own future for the sake of escaping I.C.I.'s net. (*The Observer* 11.II.1962, p. 1)

We might have expected that it would be impossible to find semantic agreement at a point to the left on the Agreement Hierarchy occurring together with syntactic agreement at a point to the right. These examples show that that expectation is false. Such examples do occur, if relatively rarely. The important claim is that at the level of the corpus the constraints of the Agreement Hierarchy will hold. That is to say, overall the likelihood of semantic agreement will be greater with targets to the right on the hierarchy but this constraint need not apply at the level of the individual sentence.

4. CONCLUSION: WHAT TO TRY NEXT?

There are several possible ways forward. The first is based on the belief that we have not really understood the syntactic structures involved. According to this view, if we had finer syntactic distinctions, then the agreement facts or at least a substantial proportion of them would fall out naturally. This could be right, but it seems unlikely. The degree of agreement variation we have seen suggests that, however fine-grained the syntax, we shall end up with many constructions which permit more than one acceptable agreement form.

A more radical proposal is that put forward by Barlow (1988), according to which there is no direct link between agreement controller and target, rather they both instigate discourse referents. The proper domain for treating agreement is discourse. His account is attractive and suggestive, but is not yet worked out in sufficient detail, with explicit analyses of the most difficult cases, to be fully convincing.[8] Pollard and Sag (1994: 72) take a related

---

[7] Pollard and Sag (1994: 87) distinguish nouns such as those in their (30) denoting social/professional castes or classes from those denoting social organizations, since *all faculty* is acceptable, while **all family* is not, but that distinction does not seem relevant to the general point being made in (30) and the data from Juul to be discussed.

[8] A specific point, relevant to our discussion of the Agreement Hierarchy, is that Barlow proposes that the mapping from morphology to properties can be sensitive to different syntactic domains. This gives a way of describing the different agreements found at different positions on the Hierarchy (Barlow 1988: 216-217). A problem which I think Barlow's account does not address as yet, is that the mapping would also have to vary according to the controller. Thus in Russian, the mapping from number morphology is different, say, for the verb and the relative pronoun, but these mappings would also differ according to whether the

ISBN: 0 08 043 438X

position, in characterizing their approach as 'a pragmatic or discourse-orientated agreement theory'; they are explicit about the mechanisms available in their theory, but when they come to the type of problem specified in this paper (under the heading of 'hybrid agreement'), they deal with them rather briefly: 'We speculate that such a competition is unstable and will be resolved as the language evolves.' (1994: 97). I suggest rather that such competition is in fact endemic in agreement systems.

Another possibility is to state the constraints at a rather different level to that which we usually consider. Johnson and Postal (1980: 20-21, 677-687) introduced 'corpus laws', which would apply to the corpus of a language rather than to individual sentences. They did not take this very far, but it is worth reviewing, given that the regularities we have discussed appear not to constrain (or not always to constrain) individual sentences. And yet, the fact that *the committee are in agreement* ... with plural agreement will be proportionally less common that *the committee ... they* would appear to be more to do with the nature of verbal agreement and anaphora than about possible languages viewed in terms of their corpora.

Perhaps then we should consider non-discrete categories. Interestingly, Shieber envisages this possibility, when considering interpretations of constraints:

> '... declarative systems based on constraints at least hold open the prospect that the discrete solution of constraints can be relaxed by incorporating continuous approximations of various sorts, for instance, probabilistic or abductive interpretation of constraints. Such possibilities are much more difficult to maintain in the context of a formalism not based on the declarative statement of grammatical information. Whether the constraint-based view of linguistic information is amenable to such a relaxation is an open question with some intriguing possibilities ...' (Shieber 1992: 16)

If we were to pursue these 'intriguing possibilities', what would be our approximations? First we would allow agreement targets to be differentially receptive to semantically motivated agreement. In other words we would represent the Agreement Hierarchy and the Predicate Hierarchy, either directly, or by reading off values from syntactic structures. And then we would allow agreement controllers to take non-discrete values. Thus in terms of the value 'plural':[9]

> politicians > committee > person

The value for 'plural' for *committee* would fall between that for *politicians* and that for *person*. We might claim that this should be read directly from 'conceptual structure', and maintain that it is just a matter of aggregation. More specifically, we could say in HPSG terms that the characteristic of nouns like *committee* is that they introduce two potential discourse referents: one an aggregate, one a nonaggregate, while *politicians* and *person* introduce only one (plural and singular respectively). This would also allow us to stay with discrete feature values. But we have failed to address the point that English *committee* readily takes plural agreement while its Russian equivalent does not and, as we saw, there are differences between British, American and New Zealand English. Moreover, other nouns like

---

controller consists of conjoined noun phrases (where the targets just mentioned allow singular or plural) or a noun phrase headed by the noun *para* 'couple' (where those targets stand in the singular, but the personal pronoun is plural), and so on. Pollard and Sag (1994: 97) would face a similar problem in that they choose between index features and anchoring conditions for the agreement of particular target types; it is not clear how these can be varied according to the controller.

[9] This recalls but is distinct from Allan's (1980) discussion of countability preferences of nouns, in which he established a scale of eight degrees of countability for English. We are considering having scalar values, 'more plural' or 'less plural' while Allan was concerned with count versus non-count.

ISBN: 0 08 043 438X

government and *party* vary from *committee* in their readiness to take plural agreement (Nixon 1970):

committee > government > party

Continuing with number, we saw how in several Slavonic languages, including Russian, there is a difference in the likelihood of numeral phrases controlling plural agreement:

2 > 3 > 4 > 5

Again we might say that the answer is to be found in conceptual structure, here it is a matter of individuation (the more individuated, the more likely is semantically justifiable agreement).[10] But then individuation must vary from language to language. If we consider a phrase like *five girls* and its translation equivalents, the likelihood of semantically justifiable agreement will vary, being most likely in English, less so in Russian, and less so again in Serbo-Croat (though here there are syntactic differences we can appeal to). And when number is used for other purposes, as we saw for honorific address, we still see variation across languages (see Table 5). It appears that honorific *vi* is 'more plural' in some languages than in others.

At least if we could limit the problem to lexical entries, we could avoid the worst of the difficulties introduced by non-discreteness. But as we saw (Table 2), the variation we are considering is not restricted to controllers headed by specific lexical items, but extends to rather general constructions such as conjoined noun phrases.

Thus language data appear to support a high degree of non-discreteness.[11] The natural response, given rather regularly, is to split agreement into two different phenomena. We have seen that this proves inadequate. However, the formal implications of non-discreteness are rather daunting. Hence while agreement used to be used as a diagnostic for work on 'more interesting' syntactic constructions, it may be better seen as a diagnostic for the adequacy of our theories.

REFERENCES

Allan, Keith. (1980). Nouns and countability. *Language* **56**.541-67.

Barlow, Michael (1988). *A situated theory of agreement.* Doctoral dissertation, Stanford, Stanford University. [Published 1992. Garland, New York.]

Barlow, Michael (1991). The Agreement Hierarchy and grammatical theory. In: *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society, February 15-18, 1991: General Session and Parasession on the Grammar of Event Structure,* (Laurel A. Sutton, Christopher Johnson and Ruth Shields (Eds)), 30-40. Berkeley Linguistics Society, University of California, Berkeley.

Bauer, Laurie (1988). Number agreement with collective nouns in New Zealand English. *Australian Journal of Linguistics* **8**.247-259.

Bayer, Sam and Mark Johnson,(1995. Features and Agreement [Online]. Available: http://xxx.lanl.gov/list/cmp-lg/9506007 [June 1995].

Bresnan, Joan and Sam A. Mchombo (1987). Topic, pronoun, and agreement in Chiche ŵa. *Language* **63**.741-82.

Comrie, Bernard (1975). Polite plurals and predicate agreement. *Language* **51**.406-18.

Corbett, Greville G. (1979). The agreement hierarchy. *Journal of Linguistics* **15**.203-24.

[10] This may be part of the explanation for the distinction between *committee, government* and *party.*

[11] See Hudson (1997) and references there, for discussion of other instances of inherent variability and other approaches to it.

ISBN: 0 08 043 438X

Corbett, Greville G. (1983). *Hierarchies, targets and controllers: agreement patterns in Slavic.* Croom Helm, London.
Corbett, Greville G. (1991). *Gender.* Cambridge University Press, Cambridge
Corbett, Greville G. (forthcoming). Morphology and agreement. To appear in: *A Handbook of Morphology* (Andrew Spencer and Arnold Zwicky (Eds)). Blackwells, Oxford.
Grevisse, Maurice (1964). *Le bon usage:Grammaire française avec des remarques sur la langue française d'aujourd'hui.* J. Duculot, Gembloux. [Eighth edition.]
Hudson, Richard (1997). Inherent variability and linguistic theory. *Cognitive Linguistics* **8**.73-108.
Johansson, Stig (1979). American and British English grammar: an elicitation experiment. *English Studies* **60**.195-215.
Johnson, David E. and Paul M. Postal (1980). *Arc Pair Grammar.* Princeton University Press, Princeton.
Juul, Arne (1972). The category of number in Modern English. *English Studies* **53**.1-23.
Kathol, Andreas (forthcoming). Agreement and the syntax-morphology interface in HPSG. In: *Readings in HPSG* (Robert Levine and Georgia Green (Eds)). Cambridge University Press, Cambridge.
Keenan, Edward (1974). The functional principle: generalizing the notion 'subject of'. In: *Papers from the Tenth Regional Meeting, Chicago Linguistic Society, April 19-21, 1974,* (Michael W. La Galy, Robert A. Fox and Anthony Bruck (Eds)), 298-309. Chicago Linguistic Society, Chicago.
Kirby, Simon (1996). Function, selection and innateness: the emergence of language universals. Unpublished PhD thesis, University of Edinburgh.
Nixon, Graham (1970). An analysis of corporate-concord situations in present-day English. Unpublished MA thesis, University of Newcastle upon Tyne.
Nixon, Graham (1972). Corporate-concord phenomena in English. *Studia Neophilologica: A Journal of Germanic and Romance Philology* **44**.120-126.
Perlmutter, David M. (1972). A note on syntactic and semantic number in English. *Linguistic Inquiry* **3**.243-246.
Pollard, Carl and Ivan A. Sag (1994). *Head-Driven Phrase Structure Grammar.* University of Chicago Press, Chicago.
Sand, Diane E. Z. (1971). Agreement of the Predicate with Quantitative Subjects in Serbo-Croatian, Ph.D. dissertation, University of Pennsylvania. Distributed by University Microfilms, Ann Arbor, 72-17,420.
Shieber, Stuart M. (1986). *An Introduction to Unification-Based Approaches to Grammar* (CSLI lecture notes 4). CSLI, Stanford University, Stanford.
Shieber, Stuart M. (1992. *Constraint-Based Grammar Formalisms: Parsing and Type Inference for Natural and Computer Languages.* MIT Press, Cambridge, MA
Suprun, A. E. (1959). *O russkix čislitel´nyx.* Kirghiz State University, Frunze.
Suprun, A. E. (1961). *Staroslavjanskie čislitel´nye.* Kirghiz State University, Frunze.
Suprun, A. E. (1963a). O soglasovanii skazuemogo s podležaščim, vključajuščim količestvennye čislitel´nye v serbo-lužickix jazykax. *Serbo-lužickij lingvističeskij sbornik*, 138-53. Moscow.
Suprun, A. E. (1963b). Zametki po sintaksisu pol´skix čislitel´nyx *Pytannja slovjans´koho movoznavstva* (L´vov) **7-8**.135-45
Suprun, A. E. (1969). *Slavjanskie čislitel´nye (stanovlenie čislitel´nyx kak osoboj časti reči).* Belorussian State University, Minsk.
Večerka, R. (1960). K sintaksisu imen čislitel´nyx v staroslavjanskom jazyke. In: *Ezikovedsko-etnografski izsledvanija v pamet na akademik Stojan Romanski* (V. Georgiev (Ed.)), 195-208. Bulgarian Academy of Sciences, Sofia.
Vincenot, C. (1975). *Essai de grammaire slovène.* Mladinska knjiga, Ljubljana.
Watson, Bruce (1979). The singularity and plurality of collective nouns: a case study. *Melbourne Working Papers in Linguistics* 5.42-49.

ISBN: 0 08 043 438X