

AGREEMENT: THE RANGE OF THE
PHENOMENON AND THE PRINCIPLES OF
THE SURREY DATABASE OF AGREEMENT¹

By GREVILLE G. CORBETT
University of Surrey

ABSTRACT

Agreement is approached from the analytical decisions required for constructing a typological database. The Surrey Database of Agreement provides detailed, highly structured information on the agreement systems of fifteen genetically diverse languages. The range of material included and the criteria for inclusion are set out here. There is then detailed discussion of the difficult cases, in particular the dividing line between agreement markers and pronominal affixes. The criteria relevant to this distinction are in part drawn from the literature and in part new. The aim is that the criteria adopted should be fully clear, so that linguists of different persuasions can use the database for their varying purposes.

¹ The research reported here was supported by the Economic and Social Research Council (United Kingdom) under grant numbers R000238228 and R00027135. This support is gratefully acknowledged. The paper takes in material from project documents by Julia Barron and Carole Tiberius, and was improved by discussion in the SMG with Matthew Baerman, Dunstan Brown, Marina Chumakina and Carole Tiberius, and especially with our project consultants Nicholas Evans and Marianne Mithun. I also wish to thank the following for helpful comments on a draft version: Julia Barron, Jonathan David Bobaljik, Kersti Börjars, Joan Bresnan, Bernard Comrie, Francis Cornish, Edith Moravcsik, Maria Polinsky, Andrew Spencer and Nigel Vincent. While one or two suggestions could not be incorporated (one came with the estimate that it would involve an additional three years of research), the comments were provided from different perspectives and have proved extremely useful.

1. INTRODUCTION

This paper tackles the question of what counts as agreement, based on the experience gained from work on the Surrey Morphology Group's Database of Agreement. Setting up this database required us to address the essential analytical issues which arise when researching the complex phenomenon of agreement. This reanalysis will serve as a basis for subsequent discussion in this volume. For practical purposes we shall also stress the underlying principles of the database, so that users can see how material has been entered, and can use this resource accordingly, whatever their theoretical persuasion.

The Surrey Database of Agreement is a novel sort of typological database, in that it includes detailed information on a small, carefully chosen set of languages. It was designed and implemented in ACCESS by Dunstan Brown and Roger Gentry. There are fifteen languages, each from a different family (Basque, Chichewa, Georgian, Hungarian, Kayardild, Mayali, Ojibwa, Palauan, Qafar, Russian, Tamil, Tsakhur, Turkana, Yimas, Yup'ik).² All phenomena normally treated under agreement are included, together with some which would count as instances of agreement for only some linguists. Thus we include agreement within the NP, agreement of the verb, and agreement of pronouns (relative and personal). The considerable size of the database is due to the fact that for each agreement phenomenon in each language we give a wealth of information wherever possible (thus besides straightforward subject-verb agreement, we include quantified subjects, subjects consisting of conjoined NPs and so on, giving the features involved, and any conditions or options). There are links to examples in each instance.

For each language included there is also a prose report, written by

² This is therefore a different enterprise compared with the extensive database compiled by Anna Siewierska. As reported in Siewierska (1999) the database had reached 272 languages, and thus allows effective validation of cross-linguistic generalisations. Our database has a much more restricted number of languages, but gives considerably more detail about the agreement system of each; it also has links to examples, so that the user can check whether our analysis meets their intended criteria. For description see Tiberius, Brown and Corbett (2002).

the researcher who entered the data for the language, giving sources, and enabling the user to see how decisions were made. Since some of the analytical decisions are difficult (we are dealing with areas where there is considerable uncertainty in the field), these reports are valuable in allowing the user to see the approach of the researcher, and to treat the data accordingly. In some instances experts on the particular languages answered questions and commented on the reports. Their help is acknowledged in the relevant report, and we take this opportunity to thank them again here.

The database is intended to ‘open doors’, to enable researchers to test hypotheses and to find data relevant to their research questions, and the sources to go further. Hence it is important that users can see both the data and the issues, rather than being presented with a forceful justification of a particular analysis.

We first consider the principles underlying the database (§2). Following ‘inclusivity’, we have included, as already noted, all the phenomena which we believe the typical linguist would expect to find (though of course there will always be other desirable information). There is one area where deciding on inclusion and coding of data is particularly difficult, namely, the question of pronominal affixes. This issue is discussed at length and our approach is justified in §3. Then we move to other difficult issues concerning non-canonical controllers in §4. The need to spell out the hard cases should not lead the reader to imagine that these are the main content of the database; there is also a great deal of information where the analytical decisions are straightforward, and so need no extended discussion.

2. PRINCIPLES UNDERLYING THE DATABASE

We start with general principles, which concern this sort of typological database, and are not specific to agreement (§2.1). Then we treat the definition of agreement (§2.2), and two logically possible approaches to the phenomenon (§2.3), before outlining our ‘canonical’ approach (§2.4).

2.1. *General principles*

2.1.1. *Inclusivity*

It is better to include data which some users may choose to omit as they work with the database, than to leave it out and so to limit users' options. We should not exclude users who happen to disagree with us.³

2.1.2. *Long-term usefulness*

Given the substantial effort going into the database (a three-year project), it should be of use in the long term. We therefore try to avoid 'indirect analyses', that is, analysing a language in a particular way by basing ourselves on other phenomena which suggest it is likely to have the phenomenon in question. Linkages (and parameters) accepted now may be revised or rejected in future; see further §3.2.2. Yet the notion of an atheoretical database is a chimera. We have to base ourselves on generally shared theoretical views, as far as we can, and to be explicit about the choices we have made (§2.1.3).

2.1.3. *Accuracy*

We have chosen to cover a restricted number of languages in order to treat them carefully and in depth. We provide links to examples, so that in each instance the user can see what lies behind the database entries. Given the detail of the distinctions we draw in the database and the variability of sources and interpretations, there will be instances of uncertainty. We draw attention to these in the prose report for the particular language, pointing out conflicting sources and approaches. It is more important to be clear about instances of doubt and debate than to simplify the presentation arbitrarily.

2.2. *Definitions of agreement*

Some terms will be helpful here.

³ Thus, for example, agreement in case is included, since some would treat it as agreement while some would not.

The term agreement commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another.

Steele (1978: 610)⁴

We call the element which determines the agreement the **controller**. The element whose form is determined by agreement is the **target**. When we indicate in what respect there is agreement, we are referring to agreement **features**. The syntactic environment in which agreement occurs is the **domain**. If we need to add further ‘provisos’, then we are dealing with **conditions**.

It is important to note that the asymmetry implied by the terms ‘controller’ and ‘target’ is a logical asymmetry, which does not prejudice how the relation should be modelled. The logical asymmetry is seen in two interrelated ways. First, the controller may have no choice of feature value, while the target does; if a verb agrees in gender with its subject, the gender of the subject NP is determined by that of the head noun, while the verb has different gender values available. And second, the contribution of the agreement features to semantic interpretation is related to the controller rather than to the target. This asymmetry can be modelled in different ways. Formerly it was done directly by copying, but there are serious problems with that approach. If we adopt a more modern approach based on unification, then the asymmetry requires something further, such as, for instance, the HPSG use of ‘anchoring’ (Pollard and Sag 1994: 60–99).

‘Concord’ is a problematic term. Many linguists treat ‘agreement’ and ‘concord’ as synonymous. Others distinguish between them, but not in ways that are mutually consistent: some treat agreement as the superordinate term, but concord can also be found as the wider term (see Corbett Forthcoming a for details). There is no useful distinction that is consistently drawn between them, hence we shall not make one. And since for those who treat them as synonyms, ‘agreement’ is increasingly the favoured term, we shall use that term exclusively. (We are careful to distinguish different domains of agreement in the database; for users who distinguish agreement and concord in this way, the relevant sections are easy to pick out.)

⁴ For detailed discussion of definitional issues see Mel’čuk (1993).

2.3. Approaches to agreement

To structure our discussion we may characterise two logical extreme views of agreement. If either could be sustained our task would be simpler.

2.3.1. The semantic/referential view

According to this view, agreement is semantic (in particular, the agreement features are semantic), and hence there is relatively little in agreement that requires special treatment. (An illustration of this position is Dowty and Jacobson 1989. See Pollard and Sag 1994: 60–99 for interesting discussion of the problems with this approach.) Take this example:

- (1) Lucinda sometimes plays the trombone.

According to the semantic/referential view, *plays* is a predicate appropriate for a semantically singular player. Example (1) is grammatical because there is semantic compatibility between subject and predicate.

A substantial proportion of the data traditionally covered under ‘agreement’ is amenable to this approach (particularly in English). However, it does not readily cover instances of grammatical gender, as in Russian:

- (2) *lampa stojal-a v uglu.* (Russian)
lamp.SG.FEM stand.PAST-SG.FEM in corner
 ‘The lamp was standing in the corner.’

In this example there is no semantic reason for *lampa* ‘lamp’ to be of feminine gender. It is important to note that non-semantic agreement extends even to anaphoric pronouns:

- (3) A: *Mama a čajnik kipjačen-yj?*
Mummy PARTICLE kettle.SG.MASC boiled.PARTICIPLE-SG.MASC
 ‘Mummy, has the kettle boiled?’
 B: *Da-a. On uže naverno čas sto-it.*
Yes. PRONOUN.SG.MASC already probably hour stand-3.SG.
 ‘Yes. It’s probably been standing for an hour.’
 (Zemskaja and Kapanadze 1978: 242)

Here *on* ‘it’ is of masculine gender because its antecedent is masculine. It is not part of the semantics of the noun *čajnik* ‘kettle’ that it is of masculine gender.

The importance of examples like (3) should be stressed. It is sometimes assumed that (morphologically free) pronouns are beyond the reach of agreement.⁵ Examples like (3), showing gender (the canonical agreement feature), suggest that agreement extends to free pronouns.⁶ Such pronouns can be both the target of agreement and a controller of agreement (thus *on* ‘it’ controls singular agreement on the verb of its clause in (3)).

2.3.2. *The syntactic view*

In this approach, agreement is a purely syntactic phenomenon, consisting of the matching of features in specified domains. According to this view, examples (1)–(3) are all grammatical because there is matching of features, which can be modelled through unification. The examples differ in that agreement features may be semantically motivated to a greater or lesser extent.

However, agreement is not a purely syntactic phenomenon, as shown by examples like this (British English):

(4) That committee are taking forever over the decision.

Here the features do not match, and the mismatch is no syntactic quirk, it is dependent on semantic factors. (The head noun *committee* is of the right semantic type, denoting an entity comprising multiple members, though this is not a sufficient condition.) Like the semantic/referential approach, the purely syntactic view accounts for a fair proportion of the data, and similarly fails to account for part of the evidence. We must accept both the semantic and the formal properties, as in Steele’s definition in §2.1. The fact that agreement choices are available in such constructions is an issue which is hard to accommodate in some theories (for discussion see Corbett 1997).

⁵ This is also of relevance below, if on that premise the case is made that other phenomena, such as pronominal affixes, are similar to such pronouns.

⁶ Of course, users who disagree can easily disregard such instances in the database. According to ‘inclusivity’ they should be available.

2.4. *A canonical approach to agreement*

If either extreme position were tenable, then we should adopt that simple view. However, neither is. Our approach will be to start from the indisputably syntactic cases, that is, the ones that require us to see agreement as a phenomenon which cannot be explained away in other terms. Taking those instances as canonical, we then work ‘outwards’ to include the less canonical cases. And the fact that syntactic agreement in gender extends even to anaphoric pronouns shows that the outer bounds of agreement extend further than many suspect.

It will be useful to introduce the notion of canonical agreement here, and it is discussed at greater length elsewhere in the volume; the original source is Corbett (Forthcoming b), where nineteen criteria for canonical agreement are discussed (they are listed in the appendix to the current paper). Consider example (5) from Russian, and its natural equivalent in Serbian/Croatian/Bosnian (6):

- (5) ty čitaš'. (Russian)
you read.2.SG
 ‘You are reading.’
- (6) čitaš. (Serbian/Croatian/Bosnian)
read.2.SG
 ‘You are reading.’

Any linguist who accepts (6) as an instance of agreement must surely also accept (5). Either both are instances of agreement, or only (5) is. There is a useful intuition behind this view; there is a redundancy in (5), which is not present in (6). These examples illustrate one of the criteria of canonical agreement:

Criterion 1: controller present > controller absent
 (where ‘>’ means ‘is more canonical than’)

It is immediately apparent that canonical is not to be associated with frequency: constructions like (6) are much more frequent, cross-linguistically, but (5) is a more canonical instance of agreement.

Now let us take an example cited as an instance of agreement by Moravcsik (1978: 337):

- (7) The man is in the room. He is old.

Some treat anaphoric pronouns as agreement targets, others do not. There are grounds for treating them as part of agreement (see the discussion of (3) above), but they are less canonical instances than those like (5):

Criterion 14: local > non-local

The issue of locality and canonicity is the focus of Polinsky (2003, this volume).

We have given just two of the criteria. As an instance of agreement which is highly canonical consider the attributive adjectives in these Spanish examples (Max Wheeler, personal communication):

- (8) un cuadro hermos-o (Spanish)
INDEF.SG.MASC picture beautiful-SG.MASC
 ‘a beautiful picture’
- (9) un-os cuadro-s hermos-os
INDEF-PL.MASC picture-PL beautiful-PL.MASC
 ‘beautiful pictures’
- (10) un-a cortina hermos-a
INDEF-SG.FEM curtain beautiful-SG.FEM
 ‘a beautiful curtain’
- (11) un-as cortina-s hermos-as
INDEF-PL.FEM curtain-PL beautiful-PL.FEM
 ‘beautiful curtains’

In these examples the controller is present, has overt features, is consistent in the agreements it takes, and its part of speech is not relevant (this last is a vacuous criterion here). The target has bound expression of agreement, with obligatory marking, doubling the marking of the noun, the marking is regular, alliterative, and productive; the target has a single controller and its part of speech is not relevant. In terms of the domain, agreement is asymmetrical (the gender of the adjective depends on that of the noun), local, and the domain is one of multiple domains. The features are lexical (for gender), they have matching values, and do not offer any choice in values. Finally there are no conditions on the agreement.

These criteria (listed in the appendix) can be thought of as defining a multi-dimensional space within which particular constructions can be located, at different distances from the canonical point where the criteria converge. Note in particular that the criteria never conflict; each defines a measure of distance from canonical agreement. They can be summarised under three principles:

Principle I: Canonical agreement is redundant rather than informative.

Principle II: Canonical agreement is syntactically simple.

Principle III: The closer the expression of agreement is to canonical (i.e. affixal) inflectional morphology, the more canonical it is as agreement.

Once we take this ‘canonical’ approach, several things fall into place. We can say that there is agreement in (4) above, but that it is less canonical than that in (2), since the features do not match. We say that local domains (2) are more canonical than non-local domains (3). This means that rather than debating whether the specification of the features of anaphoric pronouns is ‘really’ agreement or not, we can instead view it as agreement, but of a less canonical type than, say, subject–verb agreement. Issues concerned with canonical agreement are taken up in the paper by Evans (2003, this volume).

The implication for the database (§2.1.1) is that we have been inclusive, recognising that some phenomena are central to what most linguists treat as agreement while others are ‘fringe’ instances, clearly related to agreement, but treated as such by only some linguists.

3. ISSUES SPECIFIC TO PRONOMINAL AFFIXES

It is worth reiterating that given the preceding discussion there are many constructions which are clearly instances of agreement and which therefore indisputably belong in the database. Many records involve clear instances of agreement within the NP, or agreement of the verb with a core argument. This substantial section of the paper, however, is devoted to an important problematic area.

3.1. *The background*

There has been some heated debate in the literature and there is no prospect of fully satisfying all users. Again we can see two logical extreme positions, according to which all pronominal affixes count as agreement, or no pronominal affixes count as agreement. If we could sustain either view, it would be very convenient for deciding what to include in the database.

3.1.1. *The pure agreement view*

According to the first view, once there are person/number/gender features on the verb, this is an instance of verb agreement. That is, the presence of these features is a sufficient condition for agreement. Some versions of this account further assume that such features must have a local controller, and therefore insert the required NPs in the syntactic structure if they do not appear on the surface. While a theory of features is a major part of an account of agreement, one might require more than the presence of particular features to show that agreement is involved.⁷

3.1.2. *The pure pronoun view*

According to the opposite view, once an affix can be analysed as being pronominal in nature, it is an incorporated pronoun/pronominal affix and as such does not fall within agreement. That is, agreement and pronouns are quite distinct. However, as argued above (§2.3.1), we have clear instances of free pronouns showing syntactic agreement. Thus pronouns can be agreement targets (like verbs, adjectives and so on) and so deciding that a particular affix is pronominal does not thereby exclude it from consideration.

3.2. *Terms and practical considerations*

As with other parts of the debate about agreement, there are several interlocking factors here which are often not separated out. As well as the extreme cases, most of the intervening territory can be filled.

⁷ There is a much more considered version of this view; Mel'čuk discusses agreement at length and states unequivocally that polypersonal verbs show 'the most common agreement' (1993: 342–343, 365–367).

'Pronominal-affix languages' are not a homogeneous group; they vary, including in ways which affect the status of their affixes. And within a given language, pronominal affixes may not form a homogeneous group (as Baker 2002: 52 points out).

Since pronominal-affix languages cannot be neatly circumscribed, drawing a unique line for the database becomes indefensible. It is preferable to continue with a canonical/factored approach. That is, we separate out different factors (such as morphological status and referential status), and recognise that some cases are more categorial (more generally accepted or 'easier') than others. In other words, whether incorporated pronouns are near the edge of the phenomenon of agreement or the beginning of the neighbouring phenomenon is less important than seeing the connections and differences. And the varying factors observable with pronominal affixes overlap very considerably with those found in canonical agreement systems.

3.2.1. *Current usage*

We naturally take full account of current usage. If we look at recent titles of conferences and influential papers, it is clear that many would expect pronominal affixes to have a place in a database of agreement. A current example is a chapter on inflectional morphology, by Bickel and Nichols (Forthcoming). They draw a distinction between grammatical agreement and pronominal agreement, and note:

This distinction has a long tradition (but terminology varies). The idea was first introduced by DuPonceau (1819) and von Humboldt (1836) and had a veritable renaissance in the mid-eighties of the last century (cf., among others, Jelinek 1984, Mithun 1985, Van Valin 1985, Bresnan and Mchombo 1987).

Their 'pronominal agreement' is what others discuss under pronominal affixes. In whatever way we analyse the phenomenon, it seems we are duty bound to include it.⁸

⁸ Van Valin (1985) analyses Lakhota, contrasting GB and RRG approaches, and argues that Lakhota has argument affixes on the verb which function as pronouns. He makes an explicit link to Johanna Nichols' division between head-marking and dependent-marking languages, this being a clear case of head marking. Nevertheless he talks of 'verb agreement' in this context (1985: 365).

The usage of the early account of Mallinson and Blake (1981: 42–46) is still influential. They suggest that: ‘. . . in Swahili the verb can stand on its own to form a sentence, the pronominal prefixes on the verb functioning like unstressed pronouns.’ They contrast this with Germanic languages where, they say, the verb cannot stand alone and so the agreeing elements are not pronominal. Their use of terms is as follows (1981: 46): ‘. . . we will use *agreement* as the superordinate term covering the Germanic type of verb agreement and the Swahili type. We will use the term *cross-referencing agreement* or simply *cross-referencing* for the Swahili type.’⁹

In recent work in Minimalism, the operation Agree is given a major role, covering both agreement and case government (see Chomsky 2000: 101). We shall here restrict ourselves to narrow agreement, which implies a sharper notion of the matching of features, not found in government (see Evans 2003, this volume). Adopting the broader definition would blur this important distinction.

3.2.2. *The pronominal-argument parameter*

Given the difficulty of identifying pronominal affixes, it is tempting to have recourse to the ‘pronominal-argument parameter’, according to which various phenomena cluster together in languages with pronominal affixes. However, as noted in §1.1.2, we aim to avoid such ‘indirect analyses’. The point is that if we were to use suggestions concerning the pronominal-argument parameter to determine what is a pronominal affix, then our data could not contribute directly to that debate. We would wish our data to be available for assessing such claims, and not to be determined by them. In any case, Austin and Bresnan (1996) show that the

⁹ More recently, Croft uses ‘agreement’ as a cover term for various types of marking on the verb: ‘The term “agreement” is used here to describe affixes, usually expressing the category of person, usually found on the verb, indexing a participant in the verbal event, and usually allowing expression of the participant as a noun phrase argument of the verb. There are many analyses of agreement which under certain circumstances analyze the affix as pronouns (referring expressions) that happen to be bound morphemes. In §6.3.1, I will argue that all cases of agreement should be analyzed this way (see Barlow 1988). I continue to use the term “agreement” following convention, but it should be interpreted in the sense of what Barlow calls anaphoric agreement; see §6.3.1’ (2001: 139–140 fn 3).

phenomena claimed to cluster in this parameter vary independently of each other. In a similar way, ‘pro drop’ is to be understood directly in terms of the absence of referential pronouns, not indirectly by the use of other phenomena suggested as part of the ‘pro-drop parameter’ or ‘null-subject parameter’ (for which see Jaeggli and Safir 1989). After all, it is evidence from languages which regularly omit pronouns but which do not have other properties of the parameter which has led to the demise of this particular parameter.

3.3. *The theoretical issue*

Pronominal affixes (for which see Baker 1996; Evans 1999; Mithun 1999: 189–192, 202–203, 2003, this volume) are problematic, since some treat them as agreement, while others claim they are pronouns.¹⁰ Let us first ask how they resemble and differ from each of those. In terms of syntax, pronominal affixes are like pronouns, in that they can occur in clauses without any other overt NP, such that a verb with its affixes forms a complete sentence.¹¹ This can be illustrated from the Iroquoian language Tuscarora (Mithun 1999: 190):

- (12) wahrakyétkaht
 waʔ-hrak-etkaht-ʔ (Tuscarora)
FACTUAL-3.SG.AGENT/1.SG.PATIENT-chase-PERFECTIVE
 ‘He chased me.’

When we turn to their morphology, we find that they are bound to the verb, typically they are obligatory (for some this would be criterial). They often form portmanteaux combining marking of both core arguments (Evans 1999: 262), as illustrated by the

¹⁰ There can be substantial disagreement for given language families; for instance, Rice gives references on the debate within Athapaskan, and treats the relevant items herself as ‘functional items that indicate agreement and number’ (2000: 180).

¹¹ The important point here is the ability to occur without an overt NP. Pronouns (particularly in the case of clitic pronouns) may appear ‘doubling’ an NP; see, for instance, Friedman (1993: 285) on Macedonian. The essential thing at this point in the argument is that they do not require the NP.

following Tuscarora example (Marianne Mithun, personal communication):¹²

- (13) waʔkə:tkaht
 waʔ-kə-tkaht-ʔ
FACTUAL-1.SG.AGENT/2.SG.PATIENT-chase-PERFECTIVE
 ‘I chased you.’

The 1.SG.AGENT marker on its own is *-k-*, and the 2.SG.PATIENT marker is *-eθa-*. Forming portmanteaux is a rare possibility for clitics (for examples from Kapampangan, see Mithun 1994: 249, and for Catalan, see Bonet 1995). It is something which is not characteristic of free pronouns.¹³

A worthwhile next step therefore (and in harmony with our ‘canonical’ approach) is to separate out the different factors. Here this means dissociating form and function, as in Fig. 1:

<i>syntax:</i>	non-argument	argument	
<i>linguistic element:</i>	‘pure’ agreement marker	pronominal affix	free pronoun
<i>morphology:</i>	inflectional form		free form

Figure 1. The syntax and morphology of pronominal affixes.

Fig. 1 is a simplification in that it presents boundaries as being sharper than they are. In terms of syntax, there are instances where it may be arguable whether a given element represents an argument or not.¹⁴ In morphology the boundary between inflectional and free forms is

¹² The prefix glossed as 3.SG.AGENT/1.SG.PATIENT in (12) could in this instance be segmented *-hra-ak-*, though other Tuscarora prefixes are less transparent, as in (13) (Marianne Mithun, personal communication).

¹³ However, David Gil suggests a possible example: ‘In Tagalog there is a free pronoun, “kita”, whose meaning is roughly “1st person singular non-topic acting on 2nd person singular topic”. That is to say, it is a suppletive form, which occurs instead of the expected but ungrammatical sequence “ko ka”, where “ko” is “1st person singular non-topic” and “ka” is “2nd person singular topic”’ (ALT list, 12.9.2000).

¹⁴ The obvious case concerns pro-drop phenomena; if there is, say, a marker on the verb, and there is no overt subject pronoun, different analyses might accord argument status to the absent pronoun or to the present marker.

Similarly, free pronouns are typically not used with indefinite objects,¹⁸ while pronominal affixes can be (Evans 1999: 267):

- (16) Balanda bi-mey (Dulerayek dialect)
European 3.SG/3.SG.HIGHER_OBJECT.PAST-marry.PAST_PERFECTIVE
 (a) ‘She married the white man.’
 (b) ‘She married a white man.’

It appears, therefore, that pronominal affixes genuinely fall between agreement markers and free pronouns; that is, we can read the central line of Fig. 1 as expressing a scale, rather than just a convenient representation of syntactic and morphological properties. If this is so, we should ask whether there are other criteria which relate to this scale. In doing so we shall compare across languages, based on the primary means of expression, that is, contrasting pronominal affixes in a language where they are primary and free pronouns in a language where they are primary. Thus we are comparing typical agreement with typical pronominal affix and with typical free pronoun (and not, for instance, pronominal affixes with free pronouns within languages which have both).

Fig. 1 is also of some help for distinguishing two lines of argument which appear when a marker on the verb appears in different environments, specifically together with a full NP and without. Some take a ‘syntactic view’ and suggest different functions for a single marker (e.g. it is a pronoun when there is no full NP argument and an agreement marker when there is a full NP argument). This first position has been characteristic of those working in LFG (for instance in Bresnan 2001a: 146). Some take a ‘morphological view’ and suggest that if a form cannot be distinguished across environments then it should be assigned to a single category. An example of this second position is found in Georgopoulos’ account of Palauan (1991: 49).

The different assumptions here are worth bringing out. *The first position*, at its simplest, starts from a concern to understand reference. It assumes that full NPs and free pronouns are the ‘best’ referential expressions. If one of these is found it is likely to

¹⁸ Though some languages have items termed ‘indefinite pronouns’ for this purpose, such as English *one*.

be referential. If there is also a marker on the verb, which might otherwise function as a pronoun, then in this instance it cannot be doing so and is an agreement marker. (This preserves 'functional uniqueness' or its equivalent, at the cost of giving particular markers dual function.)¹⁹ *The second position*, also at its simplest, starts from a concern to understand verbal morphology. It states that markers bound to the verb should, all things being equal, be assigned to a single category. Therefore, if particular markers can function as pronouns, they *are* pronouns irrespective of other structure. Thus even if co-occurring with a full NP or free pronoun, the bound marker is still a referring pronoun. (This preserves a single function for the marker, but means that the theory of pronominal binding has to be made more complex.) Various analyses take one of the different starting points for granted, and come to conflicting conclusions from similar data. For discussion of different labellings of the Swahili object marker which depend on these different starting positions see Seidl and Dimitriadis (1997: 381).²⁰

Let us turn to a comparison of agreement markers, pronominal affixes and free pronouns.

3.3.1. Case roles

An interesting criterion for comparison concerns the number of case roles which may be indexed by the different means. In the indisputable instances of verb agreement, we normally find that just one case role can be indexed. This may be the subject, or it may be the absolutive argument. In languages claimed to have pronominal

¹⁹ It is natural to treat giving markers a dual function as a 'cost'; however, Toivonen (2000) argues that there can be positive evidence in a given language to show that this is the better approach. She considers possessive markers in Finnish. The paper is of interest here, not only because of its explicit defence of dual function, but also because it reminds us that the issue of 'dropping' arises within the NP as well as within the clause.

²⁰ Independently, Croft (2001: 226–232) contrasts two approaches to the 'dropping' problem; he suggests that the phenomena discussed all belong together ('all indexes refer' 2001: 229) and proposes a Barlow-style analysis of agreement. Cocchi (2000) discusses Bantu verbal markers and Romance clitics; she points out the similarities, including the point that 'both sets of elements carry inflectional agreement' and suggests that 'it is hardly surprising that there has been frequent disagreement among linguists as to the syntactic status of these elements' (2000: 85). Working in a Minimalist framework, she treats both as arguments of the verb.

affixes, we typically find that these index all the main arguments (two or three),²¹ as suggested, for instance, by Mithun (1986: 197).²² Free pronouns, on the other hand, can typically index all the case roles possible in a given language; in some instances this is a fair number.

3.3.2. *Referentiality*

A second criterion is the degree of referentiality. Of the three items, agreement markers have the lowest degree of referentiality (but see Barlow 1988/1992 for a more developed view, and Cornish 2000 for discussion). Nevertheless, in a language which regularly omits pronouns, one could argue that the agreement markers are partly referential.²³ Pronominal affixes frequently are referential; see, for example, Mithun (1996, 2003, this volume). Free pronouns are normally referential, though even they may have non-referential uses, however non-typical (Evans 1999: 256–257). This is a scalar criterion: for analysing verbal markers, the more referential they are, in typical uses, the stronger the case for treating them as pronominal affixes, and the greater the restrictions on referential use, the stronger the case for treating them as agreement markers.

3.3.3. *Descriptive content*

A third, related criterion is the degree of what Lyons (1977: 664) calls ‘descriptive content’. In particular, how much lexical meaning is associated with each item, as opposed to what would be called in different traditions ‘functional’ or ‘grammatical’ meaning. We have to draw the boundary in each language investigated: grammatical

²¹ As ever there are ‘in between’ cases. Thus *Tukang Besi* has been analysed as having a pronominal affix obligatorily indexing the subject; the object is optionally indexed on the verb (Donohue 1999: 123, 128).

²² One may ask whether this is sufficient to define a ‘polysynthetic language’. Evans and Sasse (2002: 3) suggest that: ‘a prototypical polysynthetic language is one in which it is possible, in a single word, to use processes of morphological composition to encode information about both the predicate and all its arguments, for all major clause types (i.e. one-, two- and three-place predicates, basic and derived), to a level of specificity allowing this word to serve alone as a free-standing utterance without reliance on context.’

²³ Ariel (2000: 213) talks of agreement markers being ‘referential (in varying degrees)’. She also states that verbal agreement markers are ‘primarily referential expressions’ (2000: 239).

meaning recurs through the system of the given language (typically described in terms of features) while descriptive content involves additional lexical specification of a semantic nature.

We can fix the two end points of our scale relatively firmly here. Starting this time with the personal pronouns, it is well known that many have descriptive content albeit to a restricted degree. For instance there are languages where they are literally ‘personal’, that is, where they can be used for reference to persons only, with some other means being required for non-persons, and this is a distinction going beyond the grammatical features (in this case, the gender feature) for the given language. For examples and references see Corbett (1991: 245–248).

At the other end of our scale, we might expect that agreement markers would have no descriptive content. After all, the agreement system is a major locus of repeated grammatical information. However, we cannot simply exclude agreement markers here. Consider these Russian examples:

(17) *vrač stojal v uglu.* (Russian)
doctor stood.MASC in corner
 ‘The doctor was standing in the corner.’

(18) *vrač stojal-a v uglu.*
doctor stood-FEM in corner
 ‘The woman doctor was standing in the corner.’

In (17) the hearer cannot be certain of the sex of the doctor. In (18) it must be a woman. The evidence is in the agreement on the verb. It appears that the agreement ‘means’ female. But this is not the case, as the next example proves:

(19) *lampa stojal-a v uglu.*
lamp stood-FEM in corner
 ‘The lamp was standing in the corner.’

Russian has formal gender assignment, so that the feminine gender includes inanimates too. While nouns denoting females are assigned to the feminine gender, the meaning ‘female’ cannot necessarily be inferred from the use of the feminine gender. Thus the Russian

agreement markers can give rise to descriptive content, in restricted contexts. This is a typical situation for agreement markers.

Given our discussion of Fig. 1, we would expect pronominal affixes to occupy a middle position, between agreement markers and free pronouns, but here we are predicting rather than reporting. A particular problem is that in several languages analysed as having pronominal affixes there is little inflection outside the verb, which makes drawing the grammatical/descriptive line difficult. We leave this open for further work and will use this criterion tentatively (see §3.4.2 III). Two pointers should be mentioned. In Iroquoian languages, in the transitive set of affixes, grammatical patients are marked only when animate (Marianne Mithun, personal communication); this can be seen as an example of pronominal affixes showing descriptive content. And second, Toivonen's discussion of possessive affixes in Finnish, though not couched in these terms, is relevant here.²⁴ She argues that certain affixes should be treated as pronominal and it is these to which additional specifications have to be added, unlike those which she analyses as agreement markers.

3.3.4. *Balance of information*

The scale of Fig. 1 leads us naturally to a discussion of the balance of information between full nominal phrases and the three items we are discussing. We need to consider the way in which information is distributed across stretches of discourse, larger than single utterances. Alternatively, to view the same issue from the hearer's perspective, we must ask how information about entities can be cumulated from such stretches. (This is based, of course, on the notion of unification, a more fruitful approach than the older views of copying of features.)

Let us take a hearer perspective and ask what can be gained from a nominal phrase compared with agreement markers, pronominal affixes and free pronouns. At one level the answer is evident: there are potentially vast numbers of nominal expressions, while in contrast the three items under discussion are limited in number (though pronominal affixes can reach impressive numbers, as in

²⁴ It is relevant in terms of comparing agreement markers with pronominal affixes, though the data concern NPs rather than verbal markers.

Dalabon, see Evans, Brown and Corbett 2001). Given the range of alternatives, there are evidently very substantially more possible nominal expressions than there are pronouns available.²⁵

The question may be posed at a more interesting level, however. We may ask what the balance of information is in respect of features which may potentially be shared (the traditional agreement features). Specifically, are number, gender and person more likely to be established from the nominal expression or from the items we are discussing?

Let us begin with agreement markers. In what we have called canonical agreement (Corbett Forthcoming b), controller and target mark the same features, and share the same number of feature values (thus if one marks number then so will the other, and the values will also be shared, so that the marking of singular/dual/plural on one implies the same possibilities on the other). However, there are many agreement systems which are not fully canonical in this way and, interestingly, the balance may be shifted in either direction. This can again be illustrated from Russian:

- (20) ja piš-u.
I.SG write-I.SG
 'I am writing.'

Here the controller marks person and number as does the target verb. The same distinctions are made in each, an instance of canonical agreement. Now compare the past tense:

- | | |
|--------------------------------|-------------------------------|
| (21) a. ja pisał. | b. ja pisał-a. |
| <i>I.SG write.PAST.SG.MASC</i> | <i>I.SG write.PAST-SG.FEM</i> |
| 'I was writing.' (man) | 'I was writing.' (woman) |

Here both the controller and the target mark number (singular/plural). The controller marks person (three values) which is not marked on the target. Conversely the target marks gender (three values) which is not a category of first- and second-person pronouns in Russian. Here, and much more generally, we find instances where the controller makes distinctions not reflected or not fully reflected

²⁵ This is of course why anaphora works in the way it does, namely, by introducing a referent through a nominal expression – one of the many – and, having established it, continuing with pronominal expressions – a choice from relatively few.

in the target, and conversely the target makes distinctions not (fully) found on the controller. In terms of cumulating information, then, the balance may be tilted either way, with more information coming from controller or from target. Impressionistically, the overall balance is roughly equal (to demonstrate this we would need to agree how different factors should be counted).

When we turn to pronominal affixes and free pronouns, however, there is a clear shift in favour of these two. That is to say, it is rare to find (morphosyntactic) feature distinctions marked on the NP which are not also found on the pronominal affixes or free pronouns (whichever the language favours). The converse is not unusual; pronominal affixes and free pronouns frequently distinguish number in languages where normal NPs do not (this is a part of the predictions which can be drawn from the Animacy Hierarchy; see Corbett 2000: 54–66 for details). And in English, the personal pronouns make distinctions in gender which are not found in the NP. The previous evidence for a scale might lead us to predict a difference between pronominal affixes and free pronouns (recall that we are comparing the major strategy *between* languages here),²⁶ but the evidence is yet to be collected to support such a hypothesis.

Note again that there is considerable variation across languages. Mithun (1998) compares Mohawk and Central Alaskan Yup'ik, both of which have obligatory affixes on the verb, marking both core arguments. However, while Mohawk nominals provide little grammatical information (few mark number), Yup'ik nominals mark three numbers.²⁷

3.3.5. *Multirepresentation versus unirepresentation*

This brings us to the last criterion showing variation across our scale, namely, 'multirepresentation'. By this we mean the co-occurrence of elements within the same clause indexed to the same referent. In

²⁶ Within a language which has pronominal affixes and free pronouns (the latter typically being restricted in use) it is fully possible for the pronominal affixes to make distinctions which the free pronouns do not.

²⁷ In this section we are concentrating on gender, number and person, but it is also worth noting here that, paralleling the difference in the marking of number, Mohawk nominals do not mark case, while Yup'ik nominals have the full inventory of cases (cf. Fig. 1).

canonical agreement we find multirepresentation: we have a controller NP, and agreement marked on one or more targets (as in (2) above). However, varying from language to language we find instances of unirepresentation (the interesting case being where the marker on the verb is the only overt indexation of the referent).²⁸

Variation of this type has often been discussed in the literature in terms of ‘doubling’ (cf. our ‘multirepresentation’) and ‘dropping’ (cf. our ‘unirepresentation’). We have introduced our terms to avoid any processual reading. There are two helpful footnotes along this line in Bresnan (2001a: 177). ‘Pronoun incorporation’ and ‘pro drop’ were used originally to indicate operations on phrase structure. In constraint-based models, however, the terms are used without any implication of movement or deletion. We, too, are using our terms in a neutral sense: multirepresentation (‘doubling’²⁹) implies two or more elements, not that having one is somehow basic, for a given language; unirepresentation (‘dropping’) implies one element, without the implication that having two is basic.³⁰

Languages are sometimes classified too rigidly according to this phenomenon. Thus Russian is generally regarded as a language in which ‘dropping’ the subject pronoun is not normal, and yet we do find some instances of it.³¹ Its South Slavonic relative Bulgarian is described as a pro-drop (null-subject) language, but this too is not an absolute phenomenon (for comparative statistics see Dončeva 1975; for statistics on varying rates of dropping in varieties of Spanish see Cameron 1993; for data showing how dropping varies according to tense in Hebrew see Ariel 2000: 236–237; for references and data on pro drop varying according to person see Meyerhoff 2000: 113–114, 136–139).

Recall that it is not possible to maintain a straightforward link between inflectional explicitness and the presence or absence of subjects. As an example of the evidence against:

²⁸ Instances where the controller cannot be overt are dealt with in §4.3.

²⁹ Specifically on clitic doubling, note Uriagereka’s comment: ‘I call doubling “maddening” with some experience of the topic’ (2000: 405).

³⁰ While the discussion here concerns the domain of the clause, multirepresentation and unirepresentation can be applied in a natural way to agreement within the NP.

³¹ We also find, in colloquial Russian, examples of the pronoun ‘doubling’ an NP; see McCoy (1998) for discussion of data earlier described by Sirotnina.

It has sometimes been suggested that subject noun phrases in Tamil are readily dropped because verbs are marked for person, number and (in third person singular) gender. That this cannot be regarded as an entirely convincing explanation is evident from a number of facts: deletion of NP subject is no less common in Tamil when there is no agreement of the sort in question – in negative and modal forms for instance; it is significant, too, that Malayalam (like Tamil a South Dravidian language), which has no person/number/gender suffixes on verbs, dispenses just as readily with subject NPs. (Asher 1996: 115–116)

There are also languages of southeast Asia lacking relevant inflectional morphology, such as Chinese, Japanese and Korean, which show regular absence of independent pronouns (Huang 2000: 57–59).

Before leaving *pro* drop, we should note Jelinek's influential paper (1984) on Warlpiri, where she claims that the clitics are the verbal arguments and that nominals are adjoined. She takes the logical next step of discussing *pro*-drop phenomena (as in Spanish) in the same terms, and suggests that Spanish has 'pronominal suffixes' and that overt subjects represent nominal adjunction. The consequence of this reasoning would be to assign languages like Russian (with agreement) and Bulgarian (with pronominal suffixes, on that analysis) to radically different types. And yet, when the agreement systems are examined in detail, considering which features are involved, and going down to the specifics of instances where different feature values are possible and the conditions which influence the choice between them, we find substantial similarities between the Slavonic languages with frequent instances of *pro* drop and those with fewer instances (Corbett 1983).³² Austin and Bresnan (1996) give a systematic comparison of Australian languages with regard to *pro* drop, and throw considerable doubt on Jelinek's analysis.

³² For discussion of the *pro*-drop parameter with reference to Slavonic see Franks (1995: 287–304). Franks does not adopt the 'pronominal affix' position, but rather treats all the family as showing agreement, with or without *pro* drop. He points out that in Slavonic '... subject pronouns are standardly omitted in exactly those languages that have full-fledged copular agreement systems' (1995: 297).

If we now turn to pronominal affixes, we again find great differences from language to language. Some show a large proportion of clauses in natural texts in which the verb's arguments are expressed solely through the verbal affixes. Mohawk represents a dramatic case: Marianne Mithun (personal communication) reports five predicates for each syntactic nominal in one sample of spoken texts.³³

When we turn to free pronouns, these are typically in complementary distribution with full NPs having the same function within the clause. Indeed theories of pronominals often trade on the assumption that a pronoun will not be coreferential with an NP within the clause. This would be covered, for example, by Binding Theory or by the LFG principle of 'functional uniqueness' (Bresnan 2001a: 145, 158).

3.3.6. *Summary: agreement markers, pronominal affixes and free pronouns*

Let us review the characteristics discussed in the preceding subsections, gathered together in Fig. 2.

	agreement marker	pronominal affix	free pronoun	heuristic (see §3.4)
case roles	1	2	all	I
referentiality	low	high	highest	II
descriptive content	low	higher	highest	III
balance of information (features versus those of NP)	roughly equal	higher	higher	IV
multirepresentation	normal	possible	largely excluded	V

Figure 2. Agreement, pronominal affixes and free pronouns (typical instances).

³³ If the count is done according to morphological form, then the ratio is much higher, since there are many instances of morphological verbs used as nominals.

The column headed ‘heuristic’ prefigures the discussion in §3.4. As we shall see, the characteristics identified here are of varying use as heuristics.

It is important to remind ourselves that we are taking typical instances here, given the variation within each type. We should also remember that we are considering only a part of agreement in this comparison, namely, agreement of the verb. Allowing for that, we do see how pronominal affixes fall between undisputed agreement affixes on the one hand, and free pronouns on the other. Given the relative nature of these differences, it is difficult to frame tests to distinguish agreement markers from pronominal affixes.³⁴

Part of the difficulty is the assumed, often unstated, linkages between characteristics which are in principle distinct. Languages vary in how likely they are to include all arguments: this is assumed to correlate with the presence of agreement morphology, but it can vary independently. The degree to which particular elements are bound morphologically varies from language to language, and does not automatically identify their syntactic status. And the referential status of pronouns does not necessarily translate directly from language to language.

Being able to draw a sharp analytical line between pronominal affixes and agreement is of importance only if one believes agreement is restricted to the clause. If, as many believe, agreement reaches beyond the clause, then pronouns agree with their antecedents (see (3) above). On this account the difference between an agreement marker and a pronominal affix is much less significant, since the pronominal affix is then also an agreement target. This can be seen if we compare the two structures (ordering is not relevant):

(22) [pronominal affix + agreement]VERB

(23) [agreement marker]VERB

For some linguists, as we noted above, the appropriate analysis would hinge on whether a full NP is present also (for some the question is whether one *may* be present). For others the structure of

³⁴ For discussion of tests proposed in the literature and the attendant difficulties see Siewierska (1999: 230–231).

the verb with its affixes will be uniform across constructions (that is, it has one structure irrespective of whether or not an NP is present).

Let us consider the task of the hearer in each case. In (22) the hearer's task is to identify the referent of the argument of the verb. This is to be done by cumulating information from the [pronominal affix + agreement] with other information, possibly supplied by an NP. As we noted earlier, the information may unify in a straightforward way, with varying amounts of information coming from the distinct sources, or there may be conflicting information, as when feature values are assigned in one instance by grammatical criteria and in the other by semantic criteria. Since the amount of information provided by the [pronominal affix + agreement] may vary, it narrows the search to a greater or lesser degree.

If we treat (23) in the same way, then the task is essentially the same. The hearer will attempt to identify the referent of the argument of the verb. This is to be done by cumulating information from the [agreement marker] with other information, possibly supplied by an NP. The information may unify in a straightforward way, with varying amounts of information coming from the distinct sources, or there may be conflicting information. Again the [agreement marker] may provide more or less information (thus a gender marker in a language with four genders may provide more information than its counterpart in a language with two genders).

Given that the tasks are essentially equivalent, we may reasonably ask whether speakers need to distinguish the two structures. There is an important point here, which picks up the earlier discussion about pro drop and its relation to agreement. We noted there that two languages may have very similar agreement systems and be pro drop (null subject) in one case and not in the other. A similar point can be made about languages with and without incorporated pronouns (pronominal affixes).³⁵ The agreement system need not be substantially different in the two types. For example, a set of tests and arguments for distinguishing incorporated pronouns has been developed by those working within LFG, and these tests have been applied particularly to Bantu languages. Several Bantu languages

³⁵ This is a distinct case, provided one does not assume that for a language to be pro drop automatically implies it has incorporated pronouns.

have been analysed as having ambiguous subject markers but incorporated pronouns for objects. Applying this approach, Demuth and Johnson (1989) show that Setawana (a dialect of Setswana) has incorporated pronouns both as the subject marker and as the object marker. This sets Setawana apart in this respect from other Bantu languages investigated to date; however, Demuth and Johnson do not point to any differences in the agreement system (for instance, in the features involved).

Having or not having incorporated pronouns is part of the balance of how far the language is head marking rather than dependent marking (Nichols 1986). While various differences follow from this distinction, the agreement system need not be greatly affected. Certain agreement domains will differ. In other respects we find agreement systems with similar properties on both sides of this typological divide.

3.4. Decisions for the database

Given that many take pronouns to agree with their antecedents, then following the notion of inclusivity we shall include them (see Corbett Forthcoming a for arguments). There is then no doubt that we should include pronominal affixes. That leaves two questions: What is the domain?, and How do we recognise them?

3.4.1. The domain for pronominal affixes

This is the easier question. We can treat pronominal affixes just as we treat anaphoric pronouns. For third persons they may have an antecedent NP (and this may be at some distance in the discourse) or they may be used deictically (indexically). First and second persons are typically used deictically.

3.4.2. Identifying pronominal affixes

For practical purposes, this proves difficult. Many who write on the topic tend to assume which language or languages have pronominal affixes and rarely give criteria for identifying them. There is also the question of tradition: languages in certain parts of the world are normally treated as having verb agreement and those in other parts as having pronominal affixes, and the distinction is not always one

of substance. Moreover, given the gradient nature of the categories there are many cases where the distinction may not be a reasonable one to draw. Detailed work has been done in LFG (Bresnan and Mchombo 1986, 1987; Bresnan 2001a: 144–179); we shall draw on that work below. However, some of the tests are quite difficult to apply to languages of different types. We shall concentrate on the more general ones.

What follows, then, is a suggested list of heuristics. Given a marker on the verb, we are looking for tests to suggest whether it is an agreement marker or a pronominal affix. In principle, each of the criteria of Fig. 2 may be applied; however, some are of more use than others.

I. Case roles

The most generally accepted instances of pronominal affixes mark the main argument positions (usually two). Typically agreement marks only one. Nevertheless there are many analyses claiming that in certain languages one marker is pronominal and the other is not. This heuristic still has some use for the instances where two argument positions are marked (this favours the pronominal affix interpretation). However, it brings the danger that we assume what we would like to demonstrate. If pronominal affixes and agreement do differ in this regard, it would be preferable to have independent tests for them, so that this interesting distribution emerged as a result rather than an assumption.

II. Referentiality

As discussed by Evans (1999: 256), if a marker is non-committal with respect to anaphoric or referential status, as is for example the obligatory third-person-singular agreement marker in English (which is able to agree with definite, indefinite and negative expressions), then it is an agreement marker. If, on the other hand, the marker unambiguously ‘refers’ to some entity then it has the status of a pronoun and is an incorporated pronoun. Use of the marker in non-referential conditions is for Evans a diagnostic of lack of full pronominal status.

This criterion is particularly applicable to third-person affixes. It is worth noting that a language may treat the persons differently; there is good evidence that third-person markers can be grammaticised after the first and second persons, and there are many languages with first- and second-person affixes but not third-person (Yuman and Siouan families). There are languages with pronominal affixes precisely for indefinites, like ‘someone’ (Mithun 1991 for Caddoan languages, and 2003, this volume, for Navajo). We should be careful to apply the heuristic only to the remainder. The inclusion of an indefinite pronoun will constrain the possibilities for the remaining affixes, but it may not follow that they will be obligatorily definite as a result.

Markers with definite reference provide an argument for pronominal status, those which are non-committal give an argument in favour of agreement.³⁶

III. Descriptive content

The tentative nature of this criterion was pointed out above. We take as our working hypothesis that the higher the descriptive content, the greater the likelihood that we are dealing with a pronoun rather than agreement. Given its vagueness, this heuristic is of limited value.

IV. Balance of information

Where there are features marked on the marker, which are not available for NPs, this is a weak argument for pronominal-affix status. Similar numbers of features leads to no prediction.

V. Complementary distribution

Probably the most useful heuristics relate to complementary distribution, which can be investigated in various ways:

³⁶ For interesting data on the connection between reference and clitic doubling in different dialects of Spanish see Uriagereka (2000).

V.i. Multirepresentation (or 'doubling')

The intuition is that the (im)possibility of a free pronoun occurring in the same clause as the marker gives an indication as to the status of a verbal marker. Note that we can apply this heuristic 'in both directions': the pronoun may block the marker and the marker may block the pronoun.

Consider first Macushi, a Carib language, with around 15,000 speakers in Brazil and Venezuela. The data are from Abbott (1991: 24–25, 101), and are discussed in Siewierska (1999: 226). In a transitive clause, the absolutive argument (with no case marker) precedes the verb, and the ergative argument follows it. This order is seen too in the markers on the verb:

- (24) *i-koneka-'pî-i-ya* (Macushi)
3.SG-make-PAST-3.SG-ERGATIVE
 'He made it.'

In (24) the verb bears a suffix *-i* corresponding to the ergative argument (transitive subject) and a prefix *i-* corresponding to the absolutive argument. We wish to ascertain the nature of these markers.

- (25) *t-ekîn* era'ma-'pî paaka esa-'ya
REFLEXIVE-pet.ABSOLUTIVE see-PAST cow owner-ERGATIVE
 'The owner of the cow saw his own pet.'

In (25) the verb bears no markers co-indexing the ergative or the absolutive argument. Once there is a free nominal or pronominal argument, there will be no marker on the verb:

- (26) **uurî-ya* *i-koneka-'pî-u-ya*
I-ERGATIVE 3.SG-make-PAST-I-ERGATIVE
 'I, I made it.'

These data strongly suggest that the markers in (24) are pronominal. Some would treat them as pronominal affixes; for others, the fact that the verb may appear without them (they do not appear where there is a free argument as in (25)) would make them clitics.³⁷

³⁷ There are less straightforward cases. There are languages where the marker may be in complementary distribution with only full nominals and optionally co-occur with pronouns (Welsh) or conversely, where the complementary distribution is with

The converse case is even more interesting, that is, where the marker is obligatory and it prevents the occurrence of the free pronoun. This situation is found in Ngalakgan, a language of the Gunwinyguan family spoken in southern Arnhem Land. Where there is a verb inflected with a first- or second-person marker, then coreferential independent pronouns cannot occur in the same intonational phrase (Baker 2002: 60–61). It seems reasonable here to interpret the marker on the verb as pronominal.³⁸ (Independent pronouns are rare in natural speech in Ngalakgan.)

For the clear cases, then, we might suggest that if a free pronoun can co-occur without problem in the same clause, then we are dealing with agreement. And if a free pronoun is not easily possible in the same clause, then we have a pronominal affix. This test works well, but for relatively few cases. The point is that the black-and-white situation of Ngalakgan is rare. Other languages allow free pronouns, but with restrictions and/or with marked effects (such as strong contrast). Once again it becomes difficult to draw the line.

We have discussed this test with regard to a free pronoun, as being the clearest case. The possibility or not of an NP headed by a noun occurring freely in the clause is a less clear test.³⁹

pronouns, not lexical NPs (Palauan). In Kichaga, the object marker must occur with free pronouns (multirepresentation) but may or may not occur with lexical NPs (Bresnan and Moshi 1990: 151–152; in the LFG analysis the object marker is seen as marking grammatical agreement when it doubles the pronoun, but as an incorporated pronoun – which shows anaphoric agreement with a topic – in the non-doubling use; when there is a lexical NP present this is treated as a dislocated topic, Bresnan 2001a: 151). Siewierska also remarks that there may be a split between common and proper nouns. A further factor is word order, where a verbal marker may co-occur with a postverbal NP, but not with a preverbal NP.

³⁸ For the third person, the independent pronouns are so rare that the situation is unclear. However, coreferential nouns do co-occur with the verbal marker; Baker takes this as *prima facie* evidence that the verbal markers act as agreement markers under those circumstances.

³⁹ To get an idea of the complex possibilities made available through multi-representation and unirepresentation, consider Larike. This is a Central Moluccan language with 8–10,000 speakers on the western tip of Ambon Island, Central Maluku, Indonesia. Central Moluccan forms part of the Central Malayo-Polynesian subgroup of Austronesian; the data are from Laidig and Laidig (1990). They write: ‘Larike verbs agree in number and person with the subject as well as with the direct object. This is accomplished with two sets of pronominal affixes . . .’ (1990: 93). Moreover: ‘where semantically and pragmatically appropriate either the subject prefixes or the object suffixes may be dropped’ (1990: 93). They do not discuss the

regularly absent). Conversely, if we were to argue that there is an incorporated pronoun here, then we would need arguments as to which component contained the pronoun, and which then agreed with it.

V.iii. Can the markers co-occur with questioned phrases?

The intuition behind this test (Bresnan and Mchombo 1987; Bresnan 2001a: 156) is that if the marker is an incorporated pronoun, it will be linked via anaphoric agreement to the ‘topic’ of the sentence. We follow Lambrecht (1994) and understand that a referent is topical in a discourse if it is taken to be a centre of current interest in the conversation, i.e., if the speaker assumes that the hearer considers it a potential locus of predication. In contrast, the ‘focus’ is the component of a pragmatically structured proposition whereby the assertion differs from the presupposition. In other words ‘topic’ is given information, ‘focus’ is new information.

The test rests on the suggestion that no argument can be both focus and topic in the same clause; it would then follow that if an element is in focus, for instance, a questioned constituent in an interrogative, and a marker is present, then the marker is an agreement marker, rather than an incorporated pronoun. This would be equivalent to the following English sentence, where *it* is equivalent to an incorporated pronoun.

(28) *What_i did he buy it_i?

Once again however, this test is problematic in that languages where the marker only co-occurs with a postverbal nominal may not have a marker with the preverbal focused element, simply because of its position rather than its discourse status. A further potential difficulty is that a possible strategy for questions involves cleft-like structures where the presupposed part is a dependent clause.

V.iv. The nature of conditions

Here we discuss instances where the marker can be affected by other conditions (where ‘conditions’ is taken to exclude normal agreement effects, that is, the straightforward matching of feature values). Work on a typology of such conditions is at an early stage.

Let us first consider morphological preconditions. For instance,

in Russian a verb agrees in gender on condition that it is in the past tense (or the conditional which is built on the past tense). A condition of this type tells us about the bound/free dimension rather than the agreement/pronominal distinction.

However, if we turn to conditions that are not morphological in nature, these can be of help as a heuristic. If such conditions affect the *value* of a feature realised by the marker, then this suggests that the marker is an agreement marker. If such conditions affect the *presence or absence* of the marker, then this suggests that the marker is a pronominal affix. It is the notion of presence/absence which makes this heuristic part of the set grouped round the notion of complementary distribution.

In Russian, given a subject consisting of a quantified NP, the verb may have a singular or a plural marker. The choice is affected by various factors. For instance, subject–predicate word order increases the chance of plural agreement over predicate–subject order, and animate subjects favour plural agreement over inanimate. These are instances of conditions; word order is not an agreement feature.

(29) vošl-o pjat' devušek. (Russian)
came.in-SG.NEUT five girl.PL.GENITIVE
 'Five girls came in.'

(30) vošl-i pjat' devušek.
came.in-PL five girl.PL.GENITIVE
 'Five girls came in.'

Since it is the value of the feature that is affected, we take this as evidence that we are dealing with agreement.

Contrast this with Chichewa. According to Bresnan and Mchombo (1986, 1987) the presence or absence of an object marker on the verb is conditioned by word order. In a canonical SVO sentence with an overt object, there is no object marker on the verb. A postverbal object can appear within the VP on their analysis only if there is no object marker on the verb. However, the word order is flexible (all orders are possible) when there is an object marker (and here they would treat an overt object as being outside the VP). The presence or absence of the object marker is conditioned

by word order, and this suggests that we are dealing with a pronominal affix rather than an agreement marker.

VI. Heuristics with limited applicability

At this point we give tests, largely derived from Bresnan and Mchombo, which are worth bearing in mind but which may have limited applicability in some cases. (For application of the Bresnan and Mchombo tests to Micronesian languages, and to a small world sample see Song 1994. For an analysis of Swahili using their tests see Keach 1995.)

VI.i. Intonation

In determining the question of complementary distribution (heuristic V), intonation may be a valuable indicator. A particular instance is that identified by Bresnan and Mchombo: in a language with SVO word order, the object may come after the verb but it may or may not be integrated into the same clause as the verb. If there is an overt object argument following a verb with an object marker, and there is an intonation break (or other tonal evidence) between the verb and the argument, then the marker may be an incorporated pronoun, and the overt NP a topic or ‘afterthought’ (antitopic):

(31) I bought-it_i, the book_i.

Otherwise, the marker is an agreement marker.

VI.ii. Idiomatic objects

Bresnan and Mchombo show that, in Chichewa, the pure idiomatic interpretation provided by an overt object is lost if a marker co-occurs, as this would effectively topicalise the idiomatic object, corresponding to the English:

(32) *He kicked it, the bucket.

This is an interesting potential test, but one which has proved hard to generalise.

VI.iii. Quantifier scope

This is allied to the issues of definiteness and questions already discussed. The presence of markers with quantifiers like *every* and

negative operators like *nobody* is inconsistent with an incorporated-pronoun interpretation. For example, in the following sentences the elements *he* and *him* are intending to mimic the behaviour of subject and object markers, respectively. Given that they cannot be referential or anaphoric, if a language has such markers they must be grammatical agreement markers.

(33) *Nobody he-likes taxes.

(34) *Taxes affect-him everybody.

A final heuristic: since first- and second-person pronouns are typically deictic, while third-person pronouns can be deictic or anaphoric, we concentrate on third-person pronouns.

4. NON-CANONICAL CONTROLLERS

We have already addressed in part the question of agreement controllers being present (multirepresentation) or absent (unirepresentation). In this section we consider the different types of non-canonical controller.⁴¹

4.1. *No overt controller (NOC)*

Here the controller is supplied by and anchored in the discourse context, that is, we are dealing with unirepresentation (the subject pronoun is absent). The understanding is that a controller could be supplied, for emphatic or topic purpose; furthermore, as the controller features change so would the target, as the following example from Italian illustrates:

(35) *Mangia la pizza.* (Italian)
eat.3.SG the pizza
 'He/she is eating pizza.'

Agreement is third-person singular. There is no overt controller (NOC); however, the addition of one is possible (29), and the absent

⁴¹ As throughout, by 'canonical agreement' we refer to instances which are unarguably agreement, i.e., where there is an overt controller and an overt target where the features of both match exactly. What is canonical need not be frequent.

argument can be seen to control agreement if its features change as in (30) where the number of the controller has changed:

- (36) Giorgio mangia la pizza.
Giorgio eat.3.SG the pizza
 ‘Giorgio is eating pizza.’
- (37) Mangiano la pizza. (cf.: Gli studenti mangiano la pizza.)
eat.3.PL the pizza (the students eat.3.PL the pizza)
 ‘They are eating pizza.’ (‘The students are eating pizza.’)

4.2. Defective controller (DC)

This is the category label for those overt controllers which lack the canonical features for controlling the target, for example, clausal or infinitival subjects:

- (38) That Tim won was surprising.
- (39) To err is human.

Here a ‘default’ third-person-singular agreement is found in the target. There are less common examples of defective controllers. For instance, Russian quantified NPs:

- (40) vošl-o pjat’ devušek. (Russian)
came.in-SG.NEUT five girl.PL.GENITIVE
 ‘Five girls came in.’

Here the numeral governs genitive plural, leaving no nominative-case-marked nominal to trigger verbal agreement. The verb takes neuter singular ‘default’ agreement, although there is the potential for variability with a semantic ‘plural’ agreement also found (see above (30), and Corbett 1983). It might be expected that a defective controller will trigger ‘default’ agreement on the target.

4.3. No possible controller (NPC)

This is the category label for those absent controllers that cannot have any surface expression, even as a topic or extra-clausal element (they may be termed ‘null expletives’ or ‘subjectless impersonals’ in

different frameworks). For example, where we have a weather verb as agreement target, but no possibility of an overt subject controlling it, as in Italian *Piove*, **Lui piove*, **Ciò piove* ‘It’s raining’, we would propose that the controller is of category ‘no possible controller’. There would appear to be a close correlation between no possible controller in one language and expletives in other languages, cf. French *Il pleut*, English *It’s raining*. We expect similar agreement forms for a defective controller and no possible controller but this is something to be investigated.

The combination of some of the heuristics discussed earlier with these possible controller types yields the following possibilities:

Controller	Target	Agreement type ⁴²	Example
NP/Pronoun	Verb	Grammatical	(41)
NOC	Verb	Ambiguous	(42)
NOC	Pronominal affix	Anaphoric	(43)
NOC/NP/Pronoun	Free pronoun	Anaphoric	(44)

(41) John sings.

(42) Cant-a. (Italian)
Sing-3.SG
 ‘He/she is singing.’

(43) Njûchi zi-ná-wá-lum-a (Chichewa)
Bees SUBJECT_MARKER-PAST-OBJECT_MARKER-bite-INDICATIVE
 ‘The bees bit them.’

(44) John arrived. **He** sat down.

CONCLUSION

We have examined an area where the data are complex, the terminology is inconsistent, and analyses can be led by unacknowledged assumptions. The fact that the same data are analysed and reanalysed by linguists in different ways should give us pause. It appears that researchers are cutting a continuum in different ways. There are interconnected strategies for identifying the referent of

⁴² The agreement type given for reference here is from Siewierska’s tripartite typology (1999: 227), which is an elaboration of that of Bresnan and Mchombo.

the arguments of verbs, and of tracking referents through discourse. These relate, to varying degrees, to the core instances of agreement. The theoretical interest in the subject matter is to see the similarities and differences in these strategies, rather than to concentrate on drawing analytical lines. We therefore adopt a canonical approach, which allows us to array the different phenomena within a typological space.

For the practical purposes of the database, this conclusion provides a consistent approach for analysing and entering data. Moreover, the language reports specify the extent to which our tests give clear results, and the database entries are linked to examples, so that users may reinterpret the data as they wish.

Our decisions are:

- We are ‘inclusive’: agreement in case and agreement of anaphoric pronouns are both included.
- As a result, some difficult phenomena cease to be an issue of inclusion or exclusion: they are included, but the question is how they should be analysed. (For instance, pronominal affixes are included, and the domain is antecedent-anaphor, as with other pronouns, thus distinguishing them from canonical verb agreement.)
- In the difficult cases, we analyse the data according to the tests given above, and give information both through the language reports and through the examples facility in the database to enable users to interpret the data according to their point of view.

Finally, the practical business of building a typological database brings theoretical issues into sharper focus; it is a useful means as well as a worthwhile end.

APPENDIX: CRITERIA FOR CANONICAL AGREEMENT

The notion of canonical agreement is outlined in §2.4. Since other papers in this volume refer to the criteria, they are listed here. Full details can be found in Corbett (Forthcoming b). Canonical agreement is the point at which the different criteria converge (there are no conflicts) and the criteria situate phenomena which extend outwards from this point (‘>’ is to be read as ‘is more canonical than’). The

criteria are listed under the five components of our account of agreement (see §2.2), namely, controllers, targets, domains, features and conditions. Comrie (2003, this volume) provides an important addition, since canonical agreement is not ‘trigger-happy’.

1. Controllers

C-1: controller present > controller absent

C-2: controller has overt features > controller has covert features

C-3: consistent controller > hybrid controller

C-4: controller’s part of speech irrelevant > relevant (given the domain)

2. Targets

C-5: bound > free

Criterion 5 can be expanded out as:

C-5’: inflectional marking (affix) > clitic > free word

C-6: obligatory > optional

C-7: regular > suppletive

C-8: alliterative > opaque

C-9: productive > sporadic

C-10: doubling > independent

C-11: target agrees with a single controller > agrees with more than one controller

Comrie’s (2003, this volume) criterion fits here as an additional criterion:

C-NEW: target has no choice of controller > target has choice of controller (is ‘trigger-happy’)

C-12: target’s part of speech irrelevant > relevant (given the domain)

3. Domains

C-13: asymmetric > symmetric

C-14: local > non-local

C-15: domain is one of set > single domain

4. Features

C-16: feature is lexical > non-lexical

C-17: features have matching values > non-matching values

C-18: no choice of feature value > choice of value

5. Conditions

C-19: no conditions > conditions

These criteria fall under three general principles, which are given in §2.4.

Surrey Morphology Group
Department of Linguistic, Cultural and International Studies
University of Surrey
Guildford
Surrey GU2 7XH
United Kingdom
Email: g.corbett@surrey.ac.uk

ABBREVIATIONS

<i>DEM</i>	demonstrative
<i>FEM</i>	feminine
<i>MASC</i>	masculine
<i>NEUT</i>	neuter
<i>PL</i>	plural
<i>SG</i>	singular
1	first person
2	second person
3	third person

REFERENCES

- ABBOTT, MIRIAM, 1991. 'Macushi', in Desmond C. Derbyshire and Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 3, Berlin: Mouton de Gruyter, 23–160.
- ANDERSON, STEPHEN R., 1992. *A-Morphous Morphology* (Cambridge Studies in Linguistics 62), Cambridge: Cambridge University Press.
- ARIEL, MIRA, 2000. 'The development of person agreement markers: from pronouns to higher accessibility markers', in Michael Barlow and Suzanne Kemmer (eds.), *Usage-Based Models of Language*, Stanford: Center for the Study of Language and Information, 197–260.
- ASHER, R. E., 1996. 'Japanese and Tamil: some typological analogies', *Asian Cultural Studies* 22 (International Christian University Publications III-A), 109–120.
- AUSTIN, PETER & BRESNAN, JOAN, 1996. 'Non-configurationality in Australian Aboriginal languages', *Natural Language and Linguistic Theory* 14, 215–268.

- BAKER, BRETT J., 2002. 'How referential is agreement? The interpretation of polysynthetic dis-agreement morphology in Ngalakgan', in Nicholas Evans and Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis* (Studia Typologica 4), Berlin: Akademie Verlag, 51–85.
- BAKER, MARK, 1996. *The Polysynthesis Parameter*, New York: Oxford University Press.
- BARLOW, MICHAEL, 1988/1992. *A Situated Theory of Agreement*, Doctoral dissertation, Stanford University, Stanford 1988, published 1992, New York: Garland.
- BICKEL, BALTHASAR & NICHOLS, JOHANNA, Forthcoming. 'Inflectional morphology', in Timothy Shopen (ed.), *Language Typology and Syntactic Description*, revised edition, Cambridge: Cambridge University Press.
- BONET, EULÀLIA, 1995. 'Feature structure of Romance clitics', *Natural Language and Linguistic Theory* 13, 607–647.
- BRESNAN, JOAN, 2001a. *Lexical-Functional Syntax*, Oxford: Blackwell.
- BRESNAN, JOAN, 2001b. 'The emergence of the unmarked pronoun', in Géraldine Legendre, Jane Grimshaw and Sten Vikner (eds.), *Optimality-Theoretic Syntax*, Cambridge, MA: MIT Press, 113–142.
- BRESNAN, JOAN & MCHOMBO, SAM, 1986. 'Grammatical and anaphoric agreement', *Papers from the Parasession on Pragmatics and Grammatical Theory at the Twenty-Second Annual Regional Meeting of the Chicago Linguistic Society*, 278–297.
- BRESNAN, JOAN & MCHOMBO, SAM, 1987. 'Topic, pronoun and agreement in Chicheŵa', *Language* 63, 741–782.
- BRESNAN, JOAN & MOSHI, LIOPA, 1990. 'Object asymmetries in comparative Bantu syntax', *Linguistic Inquiry* 21, 147–185.
- CAMERON, RICHARD, 1993. 'Ambiguous agreement, functional compensation and nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain', *Language Variation and Change* 5, 305–334.
- CHOMSKY, NOAM, 2000. 'Minimalist inquiries: the framework', in Roger Martin, David Michaels and Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Cambridge, MA: MIT Press, 89–155.
- COCCHI, GLORIA, 2000. 'Free clitics and bound affixes: Towards a unitary analysis', in Birgit Gerlach and Janet Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax* (Linguistik Aktuell 36), Amsterdam: John Benjamins, 85–119.
- COMRIE, BERNARD, 2003. 'When agreement gets trigger-happy', *Transactions of the Philological Society* 101, 313–337.
- CORBETT, GREVILLE G., 1983. *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*, London: Croom Helm.
- CORBETT, GREVILLE G., 1991. *Gender*, Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G., 1997. 'Constraints on agreement', in Bernard Caron (ed.), *Proceedings of the 16th International Congress of Linguists*, Oxford: Pergamon. CD publication, paper no. 0518.
- CORBETT, GREVILLE G., 2000. *Number*, Cambridge: Cambridge University Press.
- CORBETT, GREVILLE G., Forthcoming a. 'Agreement: Terms and boundaries', in William Griffin (ed.), *The Role of Agreement in Natural Language: Proceedings of the 2001 Texas Linguistic Society Conference, Austin, Texas, 2–4 March 2001*. Available at: <http://www.surrey.ac.uk/LIS/SMG/projects/agreement/>
- CORBETT, GREVILLE G., Forthcoming b. 'Agreement: Canonical instances and the

- extent of the phenomenon', in Geert Booij, Janet DeCesaris, Angela Ralli and Sergio Scalise (eds.), *Studies in Theoretical Morphology: Selected Papers from the Third Mediterranean Morphology Meeting, Barcelona 2001*. Available at: <http://www.surrey.ac.uk/LIS/SMG/projects/agreement/>
- CORNISH, FRANCIS, 2000. 'L'accord, l'anaphore et la référence: quelques enjeux', in Martine Coene, Walter De Mulder, Patrick Dendale and Yves D'Hulst (eds.), *Traiani Augusti vestigia pressa sequamur: Studia linguistica in honorem Lilianae Tasmowski*, Padova: Unipress, 509–533.
- CROFT, WILLIAM, 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*, Oxford: Oxford University Press.
- DEMUTH, KATHERINE & JOHNSON, MARK, 1989. 'Interaction between discourse functions and agreement in Setawana', *Journal of African Languages and Linguistics* 11, 21–35.
- DONČEVA, LILJANA, 1975. 'Nabljudenija vǎrxu kvantitativnija aspekt na nominativnite lični mestoimenija v ruski i v bǎlgarski ezik', *Bǎlgarski ezik* 25, 1, 9–20.
- DONOHUE, MARK, 1999. *A Grammar of Tukang Besi*, Berlin: Mouton de Gruyter.
- DOWTY, DAVID & JACOBSON, PAULINE, 1989. 'Agreement as a semantic phenomenon', in Joyce Powers and Kenneth de Jong (eds.), *ESCOL '88: Proceedings of the Fifth Eastern States Conference on Linguistics*, Ohio State University, Columbus, Ohio, 95–108.
- DUPONCEAU, PETER S., 1819. 'Report of the Historical and Literary Committee', *Transactions of the American Philosophical Society* 1.xvii–xlvii, Philadelphia: Abraham Small. [cited from Mithun (1985) and Bickel and Nichols (Forthcoming)].
- EVANS, NICHOLAS, 1999. 'Why argument affixes in polysynthetic languages are not pronouns: evidence from Bininj Gun-wok', *Sprachtypologie und Universalienforschung* 52, 255–281 [Slightly revised version appeared in 2002, as 'The true status of grammatical object affixes: evidence from Bininj Gun-wok', in Nicholas Evans and Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis* (Studia typologica 4), Berlin: Akademie Verlag, 15–50.]
- EVANS, NICHOLAS, 2003. 'Typologies of agreement: some problems from Kayardild', *Transactions of the Philological Society* 101, 203–234.
- EVANS, NICHOLAS, BROWN, DUNSTAN & CORBETT, GREVILLE G., 2001. 'Dalabon pronominal prefixes and the typology of syncretism: a Network Morphology analysis', in Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2000*, Dordrecht: Kluwer, 187–232.
- EVANS, NICHOLAS & SASSE, HANS-JÜRGEN, 2002. 'Introduction: problems of polysynthesis', in Nicholas Evans and Hans-Jürgen Sasse (eds.), *Problems of Polysynthesis* (Studia Typologica 4), Berlin: Akademie Verlag, 1–13.
- FRANKS, STEVEN, 1995. *Parameters of Slavic Morphosyntax*, New York: Oxford University Press.
- FRIEDMAN, VICTOR A., 1993. 'Macedonian', in Bernard Comrie and Greville G. Corbett (eds.), *The Slavonic Languages*, London: Routledge, 249–305.
- GEORGOPOULOS, CAROL P., 1991. *Syntactic Variables*, Dordrecht: Kluwer.
- GIVÓN, TALMY, 1984. *Syntax: A Functional-Typological Introduction: I*, Amsterdam: John Benjamins.
- HUANG, YAN, 2000. *Anaphora: A Cross-linguistic Approach*, Oxford: Oxford University Press.
- HUMBOLDT, WILHELM VON, 1836. *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechtes*, Berlin: Dümmler. [cited from Bickel and Nichols (Forthcoming)]

- JAEGLI, OSVALDO & SAFIR, KENNETH J., 1989. 'The Null Subject Parameter and parametric theory', in Osvaldo Jaeggli and Kenneth J. Safir (eds.), *The Null Subject Parameter* (Studies in Natural Language and Linguistic Theory 15), Dordrecht: Kluwer, 1–44.
- JELINEK, ELOISE, 1984. 'Empty categories, case and configurationality', *Natural Language and Linguistic Theory* 2, 39–76.
- KEACH, CAMILLIA N., 1995. 'Subject and object markers as agreement and pronoun incorporation in Swahili', in Akinbuyi Akinlabi (ed.), *Theoretical Approaches to African Linguistics* (Trends in African Linguistics 1), Trenton, NJ: Africa World Press, 109–116.
- LADIG, WYN D. & LAIDIG, CAROL J., 1990. 'Larlike pronouns: duals and trials in a Central Moluccan language', *Oceanic Linguistics* 29, 87–109.
- LAMBRECHT, KNUD, 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*, Cambridge: Cambridge University Press.
- LYONS, JOHN, 1977. *Semantics: II*, Cambridge: Cambridge University Press.
- MCCOY, SVETLANA, 1998. 'Individual level predicates and pronoun doubling in colloquial Russian', in Željko Bošković, Steven Franks and William Snyder (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Connecticut Meeting 1997* (Michigan Slavic Materials 43), Michigan Slavic Publications: Ann Arbor, 231–251.
- MALLINSON, GRAHAM & BLAKE, BARRY J., 1981. *Language Typology: Cross-linguistic Studies in Syntax* (North-Holland linguistic series 46), Amsterdam: North-Holland.
- MEL'ČUK, IGOR, 1993. 'Agreement, government, congruence', *Linguisticae Investigationes* 17, 307–372.
- MEYERHOFF, MIRIAM, 2000. *Constraints on Null Subjects in Bislama (Vanuatu): Social and Linguistic Factors* (Pacific Linguistics 506), Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- MILLER, PHILIP H. & SAG, IVAN A., 1997. 'French clitic movement without clitics or movement', *Natural Language and Linguistic Theory* 15, 573–639.
- MITHUN, MARIANNE, 1985. 'Disagreement: the case of pronominal affixes and nouns', in Deborah Tannen and James E. Alatis (eds.), *Georgetown Round Table on Languages and Linguistics 1985: Languages and Linguistics: The Interdependence of Theory, Data and Application*, Washington DC: Georgetown University Press, 50–66.
- MITHUN, MARIANNE, 1986. 'When zero isn't there', in Vassiliki Nikiforidou, Mary VanClay, Mary Niepokuj and Deborah Feder (eds.), *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society: February 15–17, 1986*, Berkeley, California: Berkeley Linguistics Society, University of California, 195–211.
- MITHUN, MARIANNE, 1991. 'The development of bound pronominal paradigms', in Winfred P. Lehmann and Helen-Jo Jakusz Hewitt (eds.), *Language Typology 1988: Typological Models in Reconstruction* (Current Issues in Linguistic Theory 81), Amsterdam: John Benjamins, 85–104.
- MITHUN, MARIANNE, 1994. 'The implications of ergativity for a Philippine voice system', in Barbara Fox and Paul J. Hopper (eds.), *Voice: Form and Function* (Typological Studies in Language 27), Amsterdam: John Benjamins, 247–277.

- MITHUN, MARIANNE, 1996. 'New directions in referentiality', in Barbara Fox (ed.), *Studies in Anaphora* (Typological Studies in Language 33), Amsterdam: John Benjamins, 413–435.
- MITHUN, MARIANNE, 1998. 'Challenges for inflectional description', paper at the ESRC seminar series 'Challenges for Inflectional Description', University of Essex, 1 April 1998.
- MITHUN, MARIANNE, 1999. *The Languages of Native North America*, Cambridge: Cambridge University Press.
- MITHUN, MARIANNE, 2003. 'Pronouns and agreement: the information status of pronominal affixes', *Transactions of the Philological Society* 101, 235–278.
- MORAVCSIK, EDITH A., 1978. 'Agreement', in Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.), *Universals of Human Language: IV: Syntax*, Stanford: Stanford University Press, 331–374.
- NICHOLS, JOHANNA, 1986. 'Head marking and dependent marking grammar', *Language* 62, 56–119.
- POLINSKY, MARIA 2003. 'Non-canonical agreement is canonical', *Transactions of the Philological Society* 101, 279–312.
- POLLARD, CARL & SAG, IVAN A., 1994. *Head-driven Phrase Structure Grammar*, Chicago: University of Chicago Press.
- RICE, KEREN, 2000. *Morpheme Order and Semantic Scope: Word Formation in the Athapaskan Verb*, Cambridge: Cambridge University Press.
- SEIDL, AMANDA & DIMITRIADIS, ALEXIS, 1997. 'The discourse function of object marking in Swahili', in Kora Singer, Randall Eggert and Gregory Anderson (eds.), *CLS 33: Papers from the Main Session, April 17–19, 1997*, Chicago: Chicago Linguistic Society, 373–387.
- SIEWIERSKA, ANNA, 1999. 'From anaphoric pronoun to grammatical agreement marker: why objects don't make it', in Greville G. Corbett (ed.), *Agreement* (Special issue of *Folia Linguistica* XXXIII/2), 225–251.
- SIEWIERSKA, ANNA, Forthcoming. 'On the discourse basis of person agreement marking', in T. Viryanen (ed.), *Discourse Approaches to Cognition*, Berlin: Mouton de Gruyter.
- SONG, JAE JUNG, 1994. 'The verb-object bonding principle and the pronominal system: with special reference to Nuclear Micronesian languages', *Oceanic Linguistics* 33, 517–565.
- SPENCER, ANDREW, 2000. 'Verbal clitics in Bulgarian: a Paradigm Function approach', in Birgit Gerlach and Janet Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax* (Linguistik Aktuell 36), Amsterdam: John Benjamins, 355–386.
- STEELE, SUSAN, 1978. 'Word order variation: a typological study', in Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.), *Universals of Human Language: IV: Syntax*, Stanford: Stanford University Press, 585–623.
- TIBERIUS, CAROLE, BROWN, DUNSTAN & CORBETT, GREVILLE, G. 2002. 'A typological database of agreement', in Manuel González Rodríguez and Carmen Paz Suárez Araujo (eds.), *LREC2002: Third International Conference on Language Resources and Evaluation: Proceedings: VI*, Paris: European Language Resources Association, 1843–1846.
- TOIVONEN, IDA, 2000. 'The morphosyntax of Finnish possessives', *Natural Language and Linguistic Theory* 18, 579–609.
- URIAGEREKA, JUAN, 2000. 'Doubling and possession', in Birgit Gerlach and Janet

- Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax* (Linguistik Aktuell 36), Amsterdam: John Benjamins, 405–431.
- VAN VALIN, ROBERT D., 1985. 'Case marking and the structure of the Lakhota clause', in Johanna Nichols and Anthony C. Woodbury (eds.), *Grammar Inside and Outside the Clause*, Cambridge: Cambridge University Press, 363–413.
- ZEMSKAJA, E. A. & KAPANADZE, L. A. (eds.), 1978. *Russkaja razgovornaja reč': teksty*, Moscow: Nauka.