

1 Introduction: canonical agreement

Agreement is a widespread and varied phenomenon. In some of the world's languages it is pervasive, while in others it is absent. Despite extensive research, agreement remains deeply puzzling. There was a time when it was treated mainly as a tool for researching other syntactic phenomena. Yet there has also been a tradition of recognizing it as a challenging problem in its own right. Indeed agreement presents serious problems for all our theories of syntax. It is therefore worth looking first at the reasons for the continuing interest in agreement (§1.1). Part of this comes from the way in which it involves so many components of grammar (§1.2). The terminology has become somewhat confused, so I clarify the terms I shall use (§1.3). The substantial part of this chapter lays out the canonical approach to agreement (§1.4), which will form the basis for my typology. I then outline the way in which the book is structured (§1.5), and present background information which should be of value to the reader (§1.6).

1.1 The special interest of agreement

Consider the following idea:

Hypothesis I: Grammatical information will be found only together with the lexical item to which it is relevant. (False)

This hypothesis suggests a situation which is iconic, functional, sensible and understandable. Compare *dog* and *dogs*, where number is marked in accordance with the hypothesis, or *compute* and *computed*, where tense is similarly marked. This entire book presents evidence to show that Hypothesis I is also wrong.

It is surprising that grammatical meaning can be 'displaced' (Moravcsik 1988: 90), in other words, that one word can carry the grammatical meaning relevant to another. This is what happens in agreement:

(1) Mary makes pancakes.

Here *makes* is singular because *Mary* is an individual; even if she makes pancakes frequently, the number of 'pancake making events' will not affect the agreement of the verb. The verb form tells us how many Marys there are, not how many makings there are. Thus the number information on the verb is displaced. This

displaced information, or 'information in the wrong place', is not a minor issue. Agreement affects different components of grammar, as we shall see in the next section.

1.2 The place of agreement

Take another simple example like:

- (2) The cooks make pancakes.

We need to specify that the form *make* ~ *makes* varies according to the subject (there is no effect if we change the object *pancakes* to *bread*, for example). Clearly, then, agreement is a matter of **syntax**, since the syntactic role of the items involved is of importance. But now compare:

- (3) The committee has agreed.
 (4) The committee have agreed.

Here there is a choice in some varieties of English, notably in British English. That is, there is a choice here, but not with *Mary* in (1) above. Why not? Because *Mary* is an individual, whereas *committee* may be conceptualized as an entity or as several individuals. Clearly, then, agreement is also a matter of **semantics**.

Particularly if we start from English data, we might think that agreement is all a matter of semantics, an idea put most consistently in Dowty & Jacobson (1989). We could argue that the singular verb in (1) results from semantic compatibility with a singular actor, and the plural in (2) similarly from a plural actor. However, there are three types of problem with such a view.

Consider first these examples from Morgan (1984: 235):

- (5) More than one person has failed this exam.
 (6) Fewer than two people have failed this exam.

Here we can see that the agreement of the verb depends on the grammatical number of the subject (shown by *person* versus *people*) and not on the meaning of the sentence (semantic plural in (5) and singular in (6));¹ another type of supporting example is given in §5.6.3.

There is a more general second argument that agreement cannot be entirely semantic which involves agreement in grammatical gender, in languages like Russian:

Russian

- (7) Lamp-a stoja-l-a v ugl-u
 lamp(F)-SG stand-PST-F.SG in corner-SG.LOC
 'The lamp was standing in the corner.'

¹ For the form of pronouns with such phrases see Gil (2001).

In this example there is no semantic reason for *lampa* 'lamp' to be of feminine gender.² A similar argument can be made with grammatical number in English. The use of plural agreement with English *scissors* does not, for many linguists at least, have a semantic justification.

The third argument is that even when there are semantic reasons for a particular type of agreement, the domain in which this is possible is determined by syntax. *The committee have agreed* is fine in British English (as in (4)), which suggests that *committee* takes agreement according to its meaning. And yet **these committee* is quite unacceptable. It is syntax which determines when agreement according to meaning is possible. We shall see many more examples of such mismatches in agreement in chapter 5. And evidence from acquisition also supports the syntactic basis of agreement in English (§9.3). Thus an adequate theory requires reference both to syntactic and to semantic information (Pullum 1984).

Now consider for contrast:

- (8) The committee agreed.
 (9) Mary made pancakes.
 (10) The cooks made pancakes.

Here we see no evidence of agreement. Past tense verbs in English do not show agreement. Clearly, then, agreement is a matter of **morphology** (word structure) since we require the morphology to provide the opportunity for agreement to be indicated. Indeed agreement is arguably the major interface problem between morphology and syntax, and hence appears particularly difficult when viewed from the heartland of either component.

There is a single exception to the statement about the past tense in English, namely the verb *be* which distinguishes number in the past (*was* ~ *were*). This is something that has to be stated individually for this verb, in its lexical entry. We conclude that agreement is a matter which may have to be specified in the lexicon; it is a matter of **lexicology**.

It is tempting to try to treat all such specific irregularities within the lexicon, but some apply so broadly that this approach cannot be right. Consider this example:

Russian (19th century, from Turgenev's *Nakanune* 'On the Eve', 1860)

- (11) «Mamen'ka **plač-ut**, — šepnu-l-a ona vsled uxodivš-ej
 Mother cry-3PL whisper-PST-F.SG she after leaving-F.SG.DAT
 Elene, a papen'ka **gnevaj-ut-sja** . . .»
 Elena.DAT and father be.angry-3PL-REFL

'Your mother is crying', she whispered after Elena, who was leaving,
 'and your father is angry . . .'

The speaker is a maid, talking in turn about her mistress and her master. Here the plural verbs with singular subjects indicate that the speaker is showing respect

² Dowty & Jacobson (1989: 98–101) discuss the problem of gender and attempt to meet the objection, by suggesting that a real-world property of objects is the word which is used by convention to denote that class of objects. This is hardly convincing, in my view.

for the people referred to. There are all sorts of items which could appear in this construction. They cannot be restricted to particular lexical items, rather a range of noun phrases may be involved. The generalization involves the situation: this agreement occurs when the speaker wishes to show respect (to the referents of the noun phrases agreed with). Hence agreement can be a matter of **pragmatics**.

Agreement is increasingly recognized as of interest not just for core areas of linguistics like syntax and morphology, but also more widely, in work on acquisition and in psycholinguistics, for instance, which are topics I take up in the final chapter. Given this interest from 'outside', it is particularly important that we should be talking about the same thing. Unfortunately, the terminology is muddled, and important choices in analysis are made sometimes as much by tradition as by argument. I therefore will pay attention to key terms and to the analytic choices available.

1.3 Defining terms

I have just argued for the need for clarity in terminology. What then is it that unites the examples of agreement we have considered so far? Anderson (1992: 103) points out that agreement is 'a quite intuitive notion which is nonetheless surprisingly difficult to delimit with precision'. Indeed, while several definitions have been proposed, none is fully satisfactory; see the suggestions by Keenan (1978: 167), Lehmann (1982: 203) and Lapointe (1988). There is detailed discussion of definitional issues in Mel'čuk (1993) and a formal approach can be found in Avgustinova & Uszkoreit (2003). We shall start from a suggestion by Steele:

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)

This covers the instances we have seen. The essential notion is covariance. It is not sufficient that two items happen to share properties; the sharing must be systematic, and we see this by the fact that as one element varies so will the other.

Some terms will be useful at this stage, to allow us to generalize about different types of agreement. We call the element which determines the agreement (say the subject noun phrase) the **controller**. The element whose form is determined by agreement is the **target**. The syntactic environment in which agreement occurs (the clause for instance) is the **domain** of agreement. And when we indicate in what respect there is agreement, we are referring to agreement **features**. Thus number is an agreement feature, it has the values: singular, dual, plural and so on. This is diagrammed in Figure 1.1.

Features are directly reflected in agreement. There can be other factors (like word order) which have an effect on agreement but are not directly reflected like

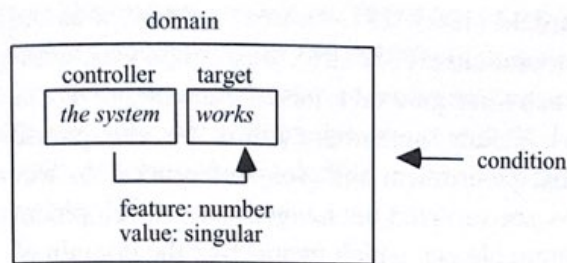


Figure 1.1 Framework of terms

features. Such factors are called agreement **conditions**. Thus, within a particular domain, a target agrees with a controller in respect of its feature specifications (that is, the features and their values); this may be dependent on some other condition being met.

These terms are now fairly standard among those working on agreement. For controller, the term 'trigger' or 'source' is sometimes found. 'Category' may be found in place of 'feature', and 'conditioning factor' for condition. For 'probe' and 'goal' see §4.2.5. As our terms suggest, there is a clear intuition that agreement is asymmetric. In *Mary laughs*, most accept that *laughs* is singular because *Mary* is singular. However, it does not follow that we should model it in this way. Older accounts of agreement captured the intuition directly by copying feature specifications from the controller to the target. More recent accounts use techniques like unification, and model the asymmetry less directly. This issue is considered in §1.4.3, and discussed more fully in §4.1.

I shall further clarify what is covered by agreement. First I deal with the term 'concord' (§1.3.1) and then I examine the relation of agreement to government (§1.3.2). The main way forward, however, will be using the notion of canonical agreement (§1.4), which will allow us to work with the full range of agreement, from the core instances of the phenomenon to those at the fringe.

1.3.1 Agreement and concord

These innocent terms have led to considerable confusion. For many linguists they are synonymous; the trend is towards the use of 'agreement',³ which is the term I shall use. Some others have distinguished the terms, but they have done so in contradictory and potentially confusing ways. Since some of this confusion remains in the literature, I shall outline two positions, so that readers can be alert to the issues. Readers for whom this is not an issue should go straight to §1.3.2.

³ In a survey of the topic, Moravcsik (1978: 333) gives 'agreement (or concord) phenomena'. Similarly Trask (1997: 10) has '**agreement** (also **concord**)'. Somewhat earlier, Lyons (1968: 239) had '*concord* (or 'agreement')'; this suggests that 'agreement' is on the rise, an impression supported by Anderson (1992: 103), who writes 'just what is "agreement" (or as it is often called in the traditional literature, "concord")?'

Some linguists, following Bloomfield (1933: 191–4), treat agreement as the superordinate term. According to Bloomfield (1933: 191), ‘In a rough way, without real boundaries, we can distinguish three general types of agreement.’ These are ‘concord’ or ‘congruence’, which includes agreement within the noun phrase and the agreement of predicate verbs, government and cross-reference.⁴ As was predictable, this system of terms has not survived unchanged. One development has been to restrict concord to the noun phrase, which means that the domain of agreement is the key part of such definitions.⁵

In contrast to the position of Bloomfield, and developments from it, Greenberg (1978: 50) treats concord as the wider term:

It would be useful, then, to distinguish the wider notion of concord from agreement, the latter being a subtype in which the choice of alternative concord elements depends on the class to which the stem of the governing item belongs, whether marked by an affix or not.

Greenberg would include matching in case within the noun phrase as an instance of concord. When, however, matching is determined by ‘the class to which the stem of the governing item belongs’, then we have agreement. Greenberg cites gender here, and is clearly talking of what we would term a lexical feature.

Note the contrast between this definition and Bloomfield’s. Most obviously the subset relations are different: for Bloomfield concord is a subset of agreement, while for Greenberg agreement is a subset of concord. But the criteria on which the relation is based differ too. Bloomfield and several followers draw a distinction according to domain: concord exists in a ‘smaller’ domain than cross-reference. For Greenberg the distinction is based on the type of feature involved: agreement involves lexical features, while concord can involve matching of other features.⁶

Thus no distinction is drawn consistently between the terms ‘agreement’ and ‘concord’, indeed they are used in opposing ways. I shall therefore use just

⁴ Bloomfield puts certain pronominal constructions and pro-drop together as cross-reference, and includes them with concord and government under ‘agreement’, but he treats antecedent-anaphor relations separately.

⁵ For example:

The term *concord* traditionally distinguishes this pattern of agreement within DP from the canonical specifier-head type: agreement theory as developed in Chomsky 1993 and related work accounts only for the latter. (Carstens 2000: 323)

Note the distinction; what for Bloomfield counted as concord is cut down to agreement within the determiner phrase, and part of what he treated as concord is treated as the ‘canonical’ type of agreement. The difference in the definitions depends on what is considered the domain of agreement.

⁶ The domain is not a defining criterion here, indeed Greenberg later talks of ‘three types of concordial phenomena’ and distinguishes in what he calls a ‘somewhat rough and heuristic fashion’ between agreement within the noun phrase, predicate agreement and anaphoric use (1978: 75–6). There are other ways in which the terms are used. Thus Lehmann (1982: 206, 249–50) also distinguishes agreement from concord: agreement is the core syntactic phenomenon, which he defines, and the term ‘concord’ is then used for instances of semantic compatibility, certain classifier-noun relations, for example. But still others use both terms without definition.

'agreement', as the more current term. There is no particular reason to determine my terms primarily according to the domains of agreement or to the features involved: both should be a part of the account (as will be the case in my 'canonical' approach). Any subdivision of agreement, whether or not 'concord' is used as the term, will require a careful definition, since there is no generally accepted terminology here.

1.3.2 Agreement and government

In the clearest instances of agreement (those I shall later treat as 'canonical'), agreement can be distinguished from government rather readily. The differences can be illustrated by this example taken from a corpus of spoken Russian.

Russian conversation (Zemskaja & Kapanadze 1978: 251)

- (12) Zna-eš' kak-oj mne vseгда dava-l-a
 know-2SG what-M.SG.ACC 1SG.DAT Always give-PST-F.SG
 sovet moj-a mam-a ?
 advice (M)[SG.ACC] my-F.SG.NOM mother(FEM)-SG
 'Do you know what advice my mother always gave me?'

The subject is *moja mama* 'my mother', and the verb agrees with it. In agreement the feature specification of the target is in the relevant respects the same as that of the controller (here feminine singular).⁷ In turn the verb governs the split noun phrase *kakoj sovet* 'what advice'.⁸ For government it is simply the presence of the verb *davat* 'give' which requires the accusative case for this noun phrase; changing the form of the verb to, say, the present, does not affect its government requirement (this is point 1 in (13) below). Another way of expressing this is to say that the agreement controller has the feature specification required of the target (i.e. the subject is indeed feminine and singular in my example), while the governor does not (the verb is not accusative), as in point 2 below. The controller of agreement is usually nominal, while targets are of various sorts; conversely, the governor can be varied, but items which are governed are nominal (point 3). The features involved in agreement, typically gender, number and person have direct semantic relevance, to varying degrees (discussed further in §4.2.4), while government canonically involves case, which is not directly involved in semantic interpretation (point 4). And finally, if there are multiple targets for an agreement controller, they will in the canonical instance share the same values (when they realize the same features); thus *moja* 'my' and *davala* 'gave' are both feminine singular.⁹ However, when a single governor governs two governees, they will

⁷ This is a further important aspect of 'systematic covariance' in the definition above, namely that it is in respect of the same feature. Thus if the case of an argument varies according to the aspect of the verb, this would not qualify as agreement any more than does normal government (thanks to Atle Grønn for pointing out this issue).

⁸ We discuss the glossing of phrases like this showing syncretism in §1.6.3 below.

⁹ As we shall see in §1.4.4, hybrid controllers are non-canonical in this regard.

normally have different feature values; thus the noun phrase *kakoj sovet* 'what advice' is accusative, while *mne* 'to me' is dative, as in point 5.

(13) Summary of differences: canonical agreement and canonical government

	AGREEMENT	GOVERNMENT
1. feature specification of target/governee is determined by:	feature specification of controller	presence of governor
2. controller/governor:	has the relevant feature specification	does not have the relevant feature specification ¹⁰
3. element which is normally nominal:	controller	governee
4. features involved are:	gender, number, person, i.e. 'direct' features (§4.2.4)	case, i.e. an 'indirect' feature
5. multiple targets/governees are:	same as each other	different from each other

In the canonical instances agreement and government are rather different, agreement being characterized by matching, and government lacking this.¹¹ However, they share the characteristic of being syntactic relations of an asymmetric type. Indeed, in recent work in Minimalism, the operation Agree is given a major role, covering both agreement and case government (see Chomsky, 2000: 101). I shall here restrict myself to agreement in the narrower sense, retaining the sharper notion of the covariance of features, not found in government. Adopting the broader definition would blur this important distinction. While I have treated the canonical instances, there are difficult phenomena falling between these idealizations, as we shall see when we consider data from Kayardild (§4.5.2). We return to the issue of agreement in case in §4.4.1, and for 'collaborative agreement', which involves an interaction with case, see §3.3.5.

1.4 Canonical agreement

To clarify some of the conceptual problems and misunderstandings that have characterized the topic of agreement I shall adopt a 'canonical' approach. This means that I shall take definitions to their logical end point and build a

¹⁰ Strictly speaking, it does not *necessarily* have the relevant feature specification; it may have it coincidentally. For example, if we have a verb which governs the genitive, a participle formed from it may be in the genitive. The fact that this participle then governs the genitive is still a matter of it being present, and does not depend on its being in the genitive.

¹¹ For extended discussion of definitions of agreement and government see Schmidt & Leffeldt (1995).

theoretical space of possibilities. Only then do I ask how this space is populated. It follows that canonical instances, which are the best and clearest examples, those most closely matching the 'canon', may well not be the most frequent. They may indeed be extremely rare. However, they fix a point from which occurring phenomena can be calibrated. Then I discuss weakenings of the criteria, which allow for less canonical instances. As these instances no longer fully match the definitions, they will include some which not all linguists would accept as instances of agreement. At several points I introduce here interesting phenomena which are then taken up in more detail in later chapters.

To start from an instance of canonical agreement, consider agreement in gender in the Italian noun phrase:

Italian (Pierluigi Cuzzolin, personal communication)¹²

- (14) il nuov-o quadr-o
 DEF.M.SG new-M.SG picture(M)-SG
 'the new picture'
- (15) i nuov-i quadr-i
 DEF.M.PL new-M.PL picture(M)-PL
 'the new pictures'
- (16) la nuov-a tel-a
 DEF.F.SG new-F.SG painting(F)-SG
 'the new painting'
- (17) le nuov-e tel-e
 DEF.F.PL new-F.PL painting(F)-PL
 'the new paintings'

I shall discuss canonical aspects of such examples in turn. As a brief summary, the canonical aspects of these examples are as follows:

- controller: is present, has overt expression of features, and is consistent in the agreements it takes, its part of speech is not relevant (this is a vacuous criterion in (14)–(17))
- target: has bound expression of agreement, obligatory marking, doubling the marking of the noun, marking is regular, alliterative, productive; the target has a single controller and its part of speech is not relevant
- domain: agreement is asymmetric (the gender of the adjective depends on that of the noun), local, and the domain is one of multiple domains
- features: lexical (in one instance), matching values, not offering any choice in values
- conditions: no conditions

For some readers examples like (14)–(17) will seem familiar; however, it is worth reflecting on how interesting they are. Each is a clear counter-example to

¹² Glossing conventions are discussed in §1.6.3. Inherent features (§4.2.3) are given in parentheses. Thus gender is glossed with the noun stem; it is true that *-a* on the noun often implies feminine gender by the assignment rules (§4.3.1) of Italian, but this is not necessarily so, as with *poeta* 'poet' (masculine); similarly *-o* often implies masculine, but this is not always the case, as with *mano* 'hand' (feminine). This glossing may seem over-careful. However, when discussing agreement it is important to distinguish between what is inherent and what is contextual.

Hypothesis I. As we shall see, the different canonical aspects of agreement converge, so that agreement in gender of the modifier with the noun in the noun phrase is confirmed as the canonical instance. Phenomena which extend the instances 'outwards' are now grouped under the five components (Figure 1.1) of my account of agreement.

1.4.1 Controllers

Several of the criteria relate to the controller. An important one is that canonical controllers are present.

C-1: controller present > controller absent
(where '>' means 'more canonical than')

Compare these two similar examples:

Russian

(18) ty čita-eš'
2SG.NOM read-2SG
'you are reading'

Serbian/Croatian/Bosnian

(19) čit-aš
read-2SG
'you are reading'

In such sentences in Russian the controller is typically present, while in Serbian/Croatian/Bosnian¹³ typically it is not. I treat as canonical what is sometimes called 'grammatical agreement' rather than 'anaphoric agreement' (Bresnan & Mchombo 1987, Siewierska 1999, Bresnan 2001a: 151). An effect of adopting criterion 1 is that, for the construction we are discussing, the canonical type is restricted to relatively few languages, since the omission of subject pronouns (often referred to as 'pro drop') is common. It is important to stress that canonical is not necessarily what is 'normal' or 'common'. Several familiar examples of languages where pronominal subjects are normally included come from northern Europe (English and German being obvious examples).¹⁴

While discussions of 'dropping' concentrate on pronouns, I am making a more general point here: it is more canonical for any controller to be present rather than absent. For agreement of the adjective with the noun within the domain of noun phrase, it is more canonical for the noun to be present; similarly in

¹³ I use this designation for the South Slavonic varieties spoken in Bosnia-Herzegovina, Croatia, Montenegro and Serbia, since linguistically they show considerable similarity. An account of the breakup of Serbo-Croat is given in Greenberg (2004).

¹⁴ A particularly interesting less familiar example is Skou (New Guinea), which has elaborate agreement marking (to be discussed in §3.2.3 and §3.2.4) and which normally includes subject pronouns: the third person pronouns are regularly included and first and second person pronouns are present more often than not (Mark Donohue, personal communication). Siewierska (2004b: 268–70) suggests in addition: the Papuan languages Au, Ekari, Koiari and Vanimo, and the Austronesian languages Anejom, Fehan and Labu.

possessor-possessed agreement it is more canonical for the 'possessed' to be present.

C-2: controller has overt expression of agreement features > controller has covert expression of agreement features

Compare these French examples:

French

(20) elle est content-e
3SG.F be.PRS.3SG happy-F.SG
'she is happy'

(21) je suis content / content-e
1SG be.PRS.1SG happy[M.SG] / happy-F.SG
'I am happy'

In (20) the controller is overtly feminine: the pronoun *elle* 'she' contrasts with *il* 'he'. In (21) there is no distinction in the controller for gender (it is underspecified for gender). We treat examples like (20) as canonical in this respect, rather than those like (21). Another way of stating this criterion is that a canonical controller marks at least as many distinctions as the target. It does so in two respects: in terms of the number of features and in terms of their values. These examples make clear that I am comparing constructions and even particular examples in terms of canonicity: even within a given language one construction can show more canonical agreement than another.

On the basis of these criteria, and others to be considered below, a more general principle may be suggested (compare Moravcsik 1988: 90):

Principle 1: Canonical agreement is redundant rather than informative

This principle fits well with the definition of agreement in §1.3. In the French example *elle est contente* 'she is happy' the feminine feature is available from the controller (criterion 2). In *je suis content(e)* 'I am happy' it is not. Agreement in the canonical example is redundant. Similarly, English examples like *the horse is / the horses are* are more canonical than *the sheep is / are*. The situation where there is no controller present, and hence the only information about the controller is that supplied by the target, is non-canonical (though, as we noted, it is commonly found); this is the point of criterion 1.

Let us continue with other criteria relating to controllers.

C-3: consistent controller > hybrid controller

A consistent controller is one which controls a consistent agreement pattern. This is more canonical than one which controls different feature values. The notion 'consistent agreement pattern' is intuitively straightforward, but not quite so easy to define (for the details see Corbett 1991: 176–81 and §5.4.1 below). As a first characterization, a consistent agreement pattern is the set of agreements controlled by a typical regular controller. A hybrid controller, on the other hand,

takes agreements from more than one such pattern. It controls different feature values on different targets. An example can be found in Bulgarian:

Bulgarian (Osenova 2003: 666)

- (22) Negov-o Veličestv-o e došāl
 his-N.SG Majesty(N)-SG AUX.3SG come.PST[M.SG]
 'his Majesty has come'

Neuter agreement is found in the noun phrase, but masculine in the verbal predicate, and so the same controller takes different agreements according to the target. A consistent controller would take either neuter or masculine (or feminine) agreements, irrespective of the target: that is, it would have a consistent agreement pattern.

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

The idea is that given a domain, for instance, subject-predicate agreement, in the canonical case we do not need further information on the part of speech of the controller. For instance, in Russian we do not need to have different rules for a subject noun phrase headed by a noun as compared to one headed by a pronoun. Sometimes, however, the difference is substantial. A good example is Bayso, where the rules are rather different for pronouns as compared with nouns. This complex situation will be analysed in §5.9.

These two criteria fall under a second general principle:

Principle II: Canonical agreement is syntactically simple

Agreement varies from examples which can be captured by a relatively simple rule, to those which are exceptionally complex. The two criteria, C-3 and C-4, both point to agreement phenomena which can be captured by simple and general rules.

1.4.2 Targets

The largest number of criteria relate to the target. This makes sense, since it is the target which is the locus of agreement. These target criteria are often intertwined, though as we shall see they can be untangled in some systems.

I begin with the nature of the expression of agreement on the target, something I shall discuss in more detail in chapter 3. Stated in the most general terms, a major criterion is:

C-5: bound > free

We are concerned with the *expression* of agreement here. Some define agreement such that its expression must be bound to the target; for example, Källström (1993: 272). Matthews' definition of agreement also has this effect: 'Syntactic relation between words and phrases which are compatible, in a given construction, by virtue of inflections carried by at least one of them.' (1997: 12). Others are

more liberal. To discuss alternative possible stances on this, let us expand out the criterion:

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

The canonical expression of agreement is through affixes bound to the target, that is, through concatenative inflectional morphology. Let us accept this ‘anchoring’ of the hierarchy (we shall return to the means of inflectional marking below) and consider the other possibilities. Some treat certain uses of clitics as agreement. According to Halpern (1998: 105) verbal clitics ‘are often assumed to be types of inflectional affixes themselves, perhaps simply agreement markers’. On the other hand: ‘there are also several respects in which clitics are not like canonical agreement affixes’.

There seems to be no argument about inflectional marking being more canonical than the use of clitics; some consider clitics (particularly in clitic doubling constructions) to be an expression of agreement, some exclude them.¹⁵ Here is an example from the South Slavonic language Macedonian:

Macedonian (Victor Friedman 1993: 285 and personal communication)

- (23) kuće-to ja=kasa mačka-ta
 dog-DEF.N.SG 3SG.F.ACC=bite[3SG] cat-DEF.F.SG
 ‘the dog bites the cat’

In (23) the clitic *ja* ‘doubles’ the noun phrase *mačkata* ‘the cat’; I mark clitic boundaries with ‘=’. The clitic is singular and feminine, like its controller. In such examples, where the object is definite (which is an example of an agreement condition, the topic of chapter 6), there must be a doubling clitic pronoun (Friedman 1993: 285). Of course, clitics vary as to ‘how bound’ they are; verbal clitics are ‘more bound’ than second position clitics, and so are somewhat closer to being canonical agreement. I take up this issue in §3.2.3.

We should now ask whether a free word can be an expression of agreement. It is important to be clear that we are looking at the expression of agreement, not just at a potential stem or host. A predicate verb is a common target, but it acts as a stem (for inflectional marking) or a host (for a clitic), but is not itself the expression of agreement. (The distinction merges particularly easily with pronouns, where an anaphoric pronoun may function as such, and be a target for agreement, or may develop into a form which loses its anaphoric function and be considered, at least by some, to be entirely an expression of agreement: see Lehmann 1982: 234–41 for early discussion, Siewierska 1999, and §9.1 below).

Potentially convincing examples of free words as the expression of agreement are found in Daly languages of north Australia. For instance, Ngan’gityemerri

¹⁵ See Harris (2002: 110–13) for discussion. Woolford (2003) uses ‘cross-referencing’ as a general term covering referencing of arguments by clitics and by inflection, with ‘agreement’ reserved for the latter; this is a good convention, when one is concerned primarily with the domain of the clause.

(Reid 1997), a Daly language with two dialects, Ngan'gikurunggurr and Ngan'giwumirri, and with 100 speakers, 300 miles south-west of Darwin, Australia, has arguably fifteen genders. Of these, six genders have optional free-form generics/classifiers:

Ngan'gityemerri (Reid 1997: 177)

(24)	(syiri) magulfu	(syiri) marrgu
	STRIKE cylindrical.fighting.stick	STRIKE new
	'a new cylindrical fighting stick'	

Syiri is the free-form generic for weapon-like objects which have a striking type of contact. In its first use in (24) it is analogous to a classifier. In its second use it is more like an agreement marker. The repetition of this free form in the noun phrase is, according to one's point of view, an example of agreement with a free word as the expression of agreement, or else a phenomenon on the edge of agreement.¹⁶ There is strong evidence that such free-form generics can develop into agreement markers, as shown by Ngan'gityemerri, where the generics are still feeding the gender system (Reid 1997: 211–22); we return to this in §9.1.1.

C-6: obligatory > optional

Canonical agreement is marked obligatorily; optional marking is less canonical. This criterion is linked to the previous one (since inflectional marking is usually obligatory), but the two can vary independently. We find optional inflectional marking of agreement, if rarely, while less canonical types of marking are more likely than inflectional marking to be optional. An example is again Ngan'gityemerri (Reid 1997). Of the fifteen genders, nine are distinguished by the agreements found on agreement targets, such as adjectives:

(25)	a-syensyerrgimi	a=tyentyenmuy
	ANIM-white.rock.wallaby	ANIM=tame
	'a tame white rock wallaby'	

Reid argues that the marker on the head noun is a prefix, while that on the agreement target is a proclitic, on the basis of stress and assimilation processes (1997: 212–15). The important point for us is that the use of these agreement markers on targets is optional (1997: 168). (We might think the language has two different systems, based on generics and on proclitics, but this is not the case, since in some genders there is a generic available in addition to a proclitic agreement marker.) As noted earlier, like the proclitic agreement markers, the generics/classifiers are optional.

These instances of optionality of agreement are less canonical than, for example, the Italian examples (14)–(17), where agreement is obligatory. We shall meet

¹⁶ If it is agreement, we must ask what the target is. A possible answer would be that it is the additional classifier 'slot' in the noun phrase which is made available by the presence of the qualifying element.

further examples in §6.7.1. Examples are frequent when we look at clitic doubling. The closely related South Slavonic languages Macedonian and Bulgarian both have clitic doubling for objects (as well as inflectional subject agreement). Earlier we looked at Macedonian and noted that clitic doubling is obligatory under certain circumstances. In Bulgarian clitic doubling is 'generally optional' (Scatton 1993: 234). There are circumstances in which it is required, but overall it is found less than in Macedonian. Thus, if clitic doubling is included as a type of agreement, we can say that the type found in Macedonian is closer to canonical agreement than that of Bulgarian.

Let us move on to the morphology of the agreement marking. There are three relevant criteria here, which we consider in turn.

C-7: regular > suppletive

The canonical marking is by regular inflectional morphology (affixation). Perhaps surprisingly, we also find instances of agreement being expressed by suppletion.

Norwegian (Bokmål, Tore Nessel, personal communication)

(26) en lit-en bil
 one/a small-M.SG car[SG]
 'one small car'

(27) to små bil-er
 two small.PL car-PL
 'two small cars'

Here we see number agreement expressed through suppletion; for other adjectives it is expressed regularly.

This criterion is logically independent of the others, which is worth bearing in mind below when, in the discussion of domains, I ask whether the English pronouns *he/she/it/they* show agreement with their antecedent. One reason why some say this cannot be agreement, almost automatically, is that the pronouns would then show suppletive expression of gender and number. We can examine the domain question in other languages where the suppletion issue can be factored out, for instance in Russian where the third person pronoun *on/ona/ono/oni* 'he/she/it/they' is not suppletive (at least in the nominative); we continue with this point about pronouns in the discussion of (37) below.

C-8: alliterative > opaque

This criterion is related to the last but differs from it. Consider this example from Swahili; here '7' indicates the singular of the Swahili 7/8 gender (Corbett 1991: 43–49):

Swahili (Welmers 1973: 171)

(28) ki-kapu ki-kubwa ki-moja ki-lianguka
 SG-basket(7/8) 7-large 7-one 7-fell
 'one large basket fell'

Two characteristics of this type of agreement system deserve attention, and both may be found to a greater or lesser degree.

1. the agreement marker on the target is identical to a formant of the controller¹⁷

In (28) the initial *ki-* on the noun *kikapu* 'basket' is indeed identical to the marker found on various agreement targets in the sentence. But this is not invariably the case in Swahili, as we see if we look at a different gender, the 3/4 gender:

- (29) m-shale u-lianguka
 SG-nail(3/4) 3-fell
 'a nail fell'

Here the agreement marker does not match the noun prefix, and so the system is not fully alliterative. English has a particularly opaque system in this respect, in having *-s* and allomorphs as the marker of the plural on controllers, but as the marker of the singular on verb targets.¹⁸

The second characteristic of alliterative agreement is:

2. the same agreement marker is used for different agreement targets

In a fully canonical system all targets take the same form. If we have, say, an adjective, numeral and verb agreeing in gender with a given noun, the agreement marker will be identical, and there will be no variation in agreement within word classes (for example, all verbs will behave identically). In example (28) we found *ki-* on each target. Contrast this with Swahili gender 1/2 (Welmers 1973: 171):

- (30) m-tu m-moja a-likuja
 SG-person(1/2) 1-one 1-came
 'one person came'

The numeral takes an alliterative form, while the verb, with the prefixed form *a-*, does not. Again languages vary: some have identical or extremely similar agreement forms, others show considerable variation (see, for instance, the discussion of Tsakhur in §3.3.3).

Thus alliterative agreement is one pole of a scale along which agreement systems can be measured. It may be that no language has totally consistent alliterative agreement, but many Bantu languages show the phenomenon to a high degree, with systems considerably more consistent than that of Swahili. Particularly

¹⁷ This point relates back to the second criterion, according to which overt expression of agreement features on the controller is more canonical than covert expression. That criterion is concerned with the *fact* of overt marking; such marking may or may not be reflected in phonologically similar forms of the target, which is what criterion 8 is about. The second point is independent of criterion 2; the targets may have the same marker, even if this is not found on the controller.

¹⁸ For the different status of this inflection in African American Vernacular English see Poplack & Tagliamonte (1989, 1994) and Green (2002: 99–102); see particularly Godfrey & Tagliamonte (1999) for references, for an account of the origins of the system of African American English, and for the 'Northern Subject Rule' (discussed in §7.7.4). The general oddness of the various English systems is pointed out by Hudson (1999).

consistent alliterative systems are found elsewhere in the Niger-Congo group (see §3.5.1 for references).

C-9: productive marking of agreement > sporadic marking

The canonical situation is for each potential target of a given type to show agreement. Thus in Russian every verb shows agreement in number. Compare this with the Nakh-Daghestanian languages Chechen and Ingush, where only around 30% of the verbs show agreement (Bickel & Nichols forthcoming; Ingush is discussed in §3.3.3). Agreement may be much more sporadic. As an extreme case, in Kuwaa, a Kru language (a group within Niger-Congo), only one adjective retains agreement in number.¹⁹ This criterion is distinct from criterion 6 (obligatory agreement is more canonical than optional agreement), in that here we are comparing items across the lexicon, whereas for criterion 6 we assume that agreement is possible and ask whether it is then obligatory or not.

These last five target criteria we have discussed can be seen as aspects of a single principle:

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

I develop the notion of canonical inflectional morphology in §3.2. We now go on to three criteria which concern the target from a wider perspective, and which fall under the principle of syntactic simplicity. The first relates back to the earlier discussion of doubling:

C-10: target always agrees > target agrees only when controller is absent

A target shows more canonical agreement if the agreement occurs irrespective of the presence or absence of the controller. That is, the target must agree, rather than doing so only when the controller is absent. This criterion relates to and further specifies the controller criterion C-1 '*controller present > controller absent*'. We need two criteria in order to generalize both over types of controller and over types of target.

My example concerns the agreement of possessive forms in Chukchi, which are formed from nouns by suffixation (Skorik 1961: 240–1). When functioning as an attributive, such possessives can agree in number with the head noun, but they do so only rarely (this is therefore another example of optional agreement, which is less canonical than obligatory agreement, according to criterion 6). The main point here is that these forms are more likely to take the plural marker when the noun controller is absent than when it is present (Skorik 1961: 233).

¹⁹ The Kuwaa adjective is cited in Marchese (1988: 335), acknowledging a personal communication from R. Thompson. More generally on criterion 9, it might be thought that this criterion is of a different order, that any phenomenon is better illustrated by non-sporadic instances. The criterion has more weight than this, in that we find that those languages which have agreement which is canonical according to a significant number of other criteria tend to be those in which it is also expressed by productive morphology.

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

Canonically, a target has a single controller, as in examples (1)–(4). Sometimes the target may mark agreement more than once, in fact it may mark it up to four times as we shall see in §3.2.4. What is less canonical is for a single target (of whatever type) to agree simultaneously with more than one controller. An example of this is found in associative/possessive constructions in some Bantu languages. I shall take examples from Shona:

Shona (Welmers 1973: 178)

- (31) *Imbwa* *na-v-ana* *v-a-dz-o*
 dogs(9/10) and-PL-young(1/2) 2-ASSOCIATIVE-10-ASSOCIATIVE
 ‘the dogs and their pups’

The last item, the associative *-a-o*, has two slots for agreement, and agrees with both nouns. The head noun *imbwa* ‘dog(s)’ belongs to gender 9/10, it does not change for number, but its plural (class 10) agreement marker is *-dz-*, hence ‘dogs’ is intended. The associated noun *v-ana* ‘children, young’ is gender 1/2 and takes the plural (class 2) agreement marker *v-*.

And finally, in this section on targets:

C-12: target has no choice of controller > target has choice of controller (is ‘trigger-happy’)

This criterion is due to Comrie (2003). The idea is that in canonical agreement a target has just one potential controller. In some less canonical instances, in a given construction there can be different controllers (as alternatives, rather than simultaneously as in the last section). Comrie gives an example from the Nakh-Daghestanian language Tsez. The target in question is the matrix verb with a sentential complement. Example (32) shows the expected construction. The complement is treated as the controller of agreement, and so the agreement is in the default gender, gender IV (the genders are given in Roman numerals). The experiencer argument, as with most verbs of this type, stands in the dative, hence *eni-r* ‘mother-DAT’. (Following Polinsky & Comrie, for clarity the embedded complement is given in square brackets.)

Tsez (Polinsky & Comrie 1999: 116–117, Comrie 2003)

- (32) *eni-r* [*už-ā* *magalu* *b-āc’-ru-li*]
 Mother(II)-DAT boy(I)-ERG bread(III)[ABS] III-eat-PST_PTCP-NMLZ[ABS]
 r-iy-xo.
 IV-know-PRS

‘The mother knows that the boy ate the bread.’

Remarkably, however, in Tsez a matrix verb can instead agree with a nominal in the absolutive, which is inside the complement. In (33) the matrix verb has gender III agreement, marking agreement with *magalu* ‘bread’, an absolutive phrase which is within the sentential complement:

- (33) eni-r [už-ā magalu b-āc'-ru-li]
 Mother(II)-DAT boy(I)-ERG bread(III)[ABS] III-eat-PST_PTCP-NMLZ[ABS]
 b-iy-xo.
 III-know-PRS
 'The mother knows that the boy ate the bread.'

We return to this interesting construction in §2.4.7, and for the conditions on its use in §6.7.1. For now the important point is that, rather than having a single possible controller, the matrix verb has two potential controllers (or triggers) and so is 'trigger-happy'. Another example is Skou (Donohue 2003a: 486–7) where some verbs (which in any case agree with the subject) may additionally show further agreement marking for the subject or agree with the object (according to the feature values of the subject and object).

C-13: target's part of speech is irrelevant > is relevant (given the domain)

The intuition here is that it is more canonical to be able to specify targets at a high level, as a general part of a domain, rather than having to make additional stipulations for subtypes. Thus we treat it as canonical to specify, for instance, that attributive modifiers agree with their head noun. Thus when we discussed Swahili (28), we noted that attributive adjective and numeral both agreed. Being able to give a rule for attributive modifiers in general is a more canonical situation than that in a language like English where one would have to specify that certain types of attributive modifier agree while some do not (we shall meet the particularly non-canonical situation in Michif in §9.1.2). Criterion 13 differs from criterion 9 (productive marking of agreement is more canonical than sporadic marking) in that the latter operates within a part of speech (do all adjectives behave alike?), while the current criterion compares across parts of speech (do all targets of a particular type behave alike, irrespective of part of speech?). Criterion 13 for targets mirrors criterion 4 for controllers.

1.4.3 Domains

There are few criteria concerning domains, but they are substantial. We consider these criteria here, then in §2.3 we return to domains in more detail, justifying the need for domains in addition to controllers and targets, and investigating their variety.

C-14: asymmetric > symmetric

The use of the terms 'controller' and 'target', and indeed the arrow in Figure 1.1, imply that agreement is an asymmetric relation. We might treat this as a defining characteristic, or we may see it as a property of canonical agreement. If two items match for the same external reason, this is not canonical agreement. If one stands in a particular form because of the properties of the first, then this is potentially canonical agreement. An analogy may be helpful. If houses numbered 10 and 12

are both white because it has snowed on both, this is not canonical agreement. If Mrs White paints number 10 white and Mrs Green in number 12 paints her house white too, that is, potentially, canonical agreement.

Seeing agreement as an asymmetric relation fits well with the idea that agreement is essentially a matter of 'displaced' information. The logical asymmetry is seen in two interrelated ways. First, the controller may have no choice of feature value, while the target does, as in these examples:

Russian

- (34) nov-yj avtomobil'
 new-M.SG car(M)[SG]
 'a new car'
- (35) nov-aja mašin-a
 new-F.SG car(F)-SG
 'a new car'
- (36) nov-oe taksi
 new-N.SG taxi(N)²⁰
 'a new taxi'

Here we have an adjective agreeing with the head noun in gender. The adjective has different morphological forms available to match the gender of the noun, while the noun does not accommodate the adjective in any comparable way. Logically, then, the relation is asymmetric, with the adjective being controlled by the noun. Examples of the verb agreeing in person with the pronoun would make the point equally well.

The second part of the logical asymmetry of agreement concerns the contribution of the agreement features to semantic interpretation. In examples (34)–(36) gender is not based on semantics but depends on assignment based on form (§4.3.1). If, however, in place of *avtomobil'* 'car' and *mašina* 'car' we have *byk* 'bull' and *korova* 'cow', then we have semantically based gender. Yet the gender marking on the adjective does not affect the interpretation of *nov-* 'new'. The contribution to semantic interpretation is related to the controller rather than to the target. Again this points to the asymmetry of the agreement relation (see also Nichols 1985, 1986).

This is a logical asymmetry, which does not determine how the relation should be modelled. There have been different means for doing so. Formerly the asymmetry was handled directly by copying, but there are serious problems with that approach: the controller may be absent (as in pro-drop languages, example (19)); or it may be present but be underspecified, as in (21); or the feature specifications on the controller and the target may simply not match, as we shall see in §1.4.4. More modern approaches are based on unification, which does not capture the asymmetry directly, and so leads to the question of how it is to be captured. We discuss this important issue in §4.1.

²⁰ This noun is indeclinable and so does not mark number (see §5.1.1).

If we accept that agreement is canonically an asymmetric relation, that leads to the problem of agreement in case. For linguists who have a view of syntax which is based on the notion of constituency, the traditional instances of 'agreement in case' are not agreement: matching of case values within the noun phrase results from government of the whole noun phrase by an external governor (see (13)). For those who accept a dependency view of syntax, the opposite conclusion follows, namely that there is agreement in case. I conclude that canonical agreement is asymmetric. Which instances count as asymmetric, and therefore potentially canonical, depends on other assumptions about syntax. We consider the agreement features in chapter 4, and we look specifically at the question of agreement in case in §4.4.1.

C-15: local domain > non-local domain

This criterion implies that the 'smaller' the domain the more canonical it is. That is, the smaller the structural distance between controller and target the more canonical is the instance of agreement. The most canonical is agreement within the phrase, as in examples like *these books*, and in (14)–(17); some would call this 'concord' (§1.3.1). Less canonical would be agreement beyond the phrase but within the clause, as in *Mary sings*, showing agreement of the verb with one of its arguments. Then we have agreement beyond the clause but within the sentence; this would be agreement of the relative pronoun with its antecedent (which we meet in §2.2.2). Finally we have the more controversial domain which goes potentially beyond the sentence, namely agreement of the anaphoric (personal) pronoun with its antecedent, as in *Mary sings because she is happy*.

The question as to whether agreement is only a local phenomenon is rarely asked. Opposing views are stated, almost as facts, with little discussion. There is a divide here, though by no means an absolute one, between those who have treated agreement as a prime focus of study as opposed to those who come to it as one of a set of syntactic phenomena to be accounted for. The former, for instance Moravcsik (1978: 334) and Lehmann (1982: 211), typically assume that the feature values of anaphoric pronouns are determined by agreement mechanisms. They cite examples of anaphoric pronouns within the discussion of agreement. On the other side, those who come to agreement as just one syntactic phenomenon of many often assume that it is a local phenomenon, and so exclude examples like (*Mary . . . she*). This is a convenient delineation for syntax, but we shall see evidence to question it. The only extended discussion of the issue of which I am aware is found in Barlow (1991, 1992: 134–52), who concludes that there are no good grounds for distinguishing between agreement and antecedent-anaphor relations. Agreement cannot be restricted only to local domains. This conclusion is confirmed in Siewierska (1999: 225).²¹

There are two main types of evidence supporting this conclusion: the type of features involved, and the distribution of syntactic and semantic agreement. The

²¹ For the agreeing pronouns of Fula, which show special patterns, see Culy (1996).

simple argument is that canonical agreement and antecedent-anaphor relations are often based on the same features. This can be illustrated from a Russian example, from the transcript of a conversation:

- Russian (Zemskaja & Kapanadze 1978: 242)
- (37) Mama a čajnik kipjačen-yj?
 Mummy PARTICLE kettle(M)[SG] boiled.PST.PTCP.PASS-M.SG
 'Mummy has the kettle boiled?'
 Da-a. On uže naverno čas sto-it.
 Yes. 3[M.SG.NOM] already probably hour stand-3SG
 'Yes. It's probably been standing for an hour.'

The anaphoric pronoun *on* is masculine singular, because those are the feature values of its antecedent *čajnik* 'kettle'. Here the participle *kipjačenyj* 'boiled', like an adjective, distinguishes number (two values) and gender (three values: masculine, feminine and neuter, but only in the singular). The anaphoric pronoun does the same. It is not always the case, cross-linguistically, but it is extremely common that the anaphoric pronoun has the same feature possibilities as other agreement targets. If agreement and antecedent-anaphor relations are split, then there are two distinct phenomena which for no principled reason utilize identical features.

The second argument must wait until additional concepts have been introduced, so we will only preview it here. The four domains mentioned above constitute the Agreement Hierarchy, which will be discussed extensively in chapter 7. The hierarchy constrains the distribution of syntactic and semantic agreement. This distribution is a gradient phenomenon, across the range of domains. Evidence from the Agreement Hierarchy shows that there is no one point at which agreement phenomena can be neatly divided into two in a principled way. Rather there are several different domains for agreement, related in hierarchical fashion.

Anticipating the discussion in §2.2.2 and §7.6.1, I conclude that agreement covers feature covariance in a range of domains, from within the noun phrase to antecedent-anaphor relations. This is accepted in Head-Driven Phrase Structure Grammar, HPSG (Pollard & Sag 1994: 74), and in Lexical-Functional Grammar, LFG (Bresnan 2001a: 151). And as we shall see in §9.4.6, there is some psycholinguistic evidence to support this conclusion. Others limit agreement, more or less drastically. If we are to draw a boundary, then we need to be clear whether this is based on evidence from agreement itself (which would be hard to justify), or whether the boundary is being drawn as a result of other considerations within the syntactic model adopted. If such a boundary is proposed, then we should ask whether it claims to handle the distribution of syntactic versus semantic agreement (again such a claim seems unlikely to be well founded). However, even if one excludes antecedent-anaphor relations as part of agreement, this is likely to be because they are not local links, thus taking criterion 15 as categorical rather than gradient. Within the domains there are other sources of considerable variety. We discuss these in §2.3; the interesting issue of 'long-distance' agreement

(a term suggesting that controller and target are more distant syntactically than we would expect) will be taken up in §2.4.7.

If we accept anaphoric pronouns as agreement targets, treating antecedent-anaphor as a domain, it is worth noting that an anaphoric pronoun is a pronoun which also agrees. Since I am using the criteria to separate out overlapping factors, I have concentrated in this section on the syntactic position of such pronouns. However, their morphology can also vary, and in part independently of their syntax. Thus anaphoric pronouns can be morphologically free or bound, the latter often being termed 'pronominal affixes' or 'incorporated pronouns' (discussed in §3.8). Pronominal affixes are less canonical in terms of their domain than, say, subject-verb agreement, since they are part of a non-local domain; on the other hand, they are more canonical than free pronouns in being morphologically bound. It is generally accepted that diachronically pronouns provide a major source of agreement morphology, progressing from full pronouns, to clitics, to inflections, as we shall see in §9.1.

Finally in this section on domains we shift from looking at individual relations to looking at the system, hence our last criterion is couched in terms of a given domain (and its being one of several).

C-16: domain is one of a set > single domain

In canonical instances, a given domain will be a member of a set of domains (agreement with a given controller may be expressed by different targets), following a general rather than a specific syntactic rule. Thus if we take Russian subject-verb agreement, this is one domain of several (attributive modifier agreeing with head noun, relative with antecedent . . .). This is a more canonical situation than that in a language where, say, subject-verb is the only agreement domain.

This criterion links back to the notion of redundancy: information concerning a given controller can be expressed more than once in different domains. An interesting implication related to this criterion is that multiple domains may well be a sufficient but not necessary condition for showing that particular markers are agreement markers rather than pronominal affixes (incorporated pronouns). Where different targets can show what is claimed to be agreement with a single controller, it is much more likely that these are instances of agreement rather than being pronominal affixes (§3.8.2).

1.4.4 Features

Here we find three criteria, one relating to features as a whole, and two relating to their values. Features are discussed in detail in chapter 4.

C-17: feature is lexical > non-lexical

Agreement in gender (where lexical) is considered the canonical type (see further §4.2.3, where I show that lexical features are the core of the 'inherent' features). The reason is that the target could not be marked with the feature independently,

if it is lexical, and so this links to the asymmetry of agreement. Thus in (34)–(36) there is no independent source of the gender feature apart from the controller. Another way of stating this criterion is that features which are based at least in part on formal assignment are more canonical for agreement than features where assignment is more semantically based (§4.3.1). This criterion therefore falls under the principle of redundancy.

An interesting consequence concerns anaphoric pronouns; the fact that in many languages these can covary according to lexical gender strongly suggests they are part of the phenomenon of agreement, as discussed in relation to (37).

C-18: features have matching values > non-matching values

This seems obvious: some would claim that the definition of agreement must refer to the matching of values (§1.3). However, once a construction is identified as involving agreement, because there is a covariance of features, we would not want to rule out the analogous instances where the features do not match.²² Specifically, since English subject and predicate verb regularly have matching features, we have to address examples like this one where they do not:

(38) the committee have decided

We cannot simply say that *committee* is plural, since we find *this committee* and not **these committee*. We need to invoke a notion of semantic agreement for such cases, that is, agreement consistent with the meaning of the controller (discussed in detail in §5.4). From this point of view, we can say that examples like (38) are less canonical instances of agreement than those where the feature values match straightforwardly (*the committee has decided*); for further discussion see Corbett (2000: 188–91). Mismatches are analysed in chapters 5 and 7.

If we accept that semantic agreement is non-canonical, then we should include here instances of resolution, which specifies the feature values of targets when the controller consists of conjoined noun phrases. Consider this example from Slovene (Priestly 1993: 433):

Slovene (Priestly 1993: 433)

(39) Milk-a in njen-o tele sta bi-l-a zunaj
 Milka(F)-SG and her-N.SG calf(N)[SG] AUX.3DU be-PST-M.DU outside
 'Milka and her calf were outside.'

Here we have a feminine singular and a neuter singular conjoined; the verb is dual and masculine. Clearly, then, the features do not match. It is resolution which specifies these particular feature values (as we shall see in chapter 8). The fact that such instances are taken to be non-canonical fits with §8.6, where the peripheral nature of resolution rules is discussed.

The general effect of this criterion is to claim that syntactic (formal) agreement is more canonical than semantic agreement. An interesting consequence is

²² Mel'čuk (1993: 329–31) stresses that the definition of agreement must allow for such instances; in Steele's definition this is covered by the reference to a semantic property of the controller.

that unification is an adequate mechanism for formalizing canonical instances of agreement (discussed further in §4.1). This consequence demonstrates well that criterion 18 falls under the principle of syntactic simplicity. The criterion is also consonant with the 'redundancy' principle.

Non-matching values can arise in various circumstances, from those which can be related directly to the lexical item (as in (38)), through those involving a construction (39), to those which depend on the use of the item, the pragmatics (as in (11)). There are systems in which mismatching is widespread, systems which Bickel (2000) calls 'associative', which are less canonical than the more familiar 'integrative' systems (of languages like Russian). We return to mismatches in chapter 5.

C-19: no choice of feature value > choice of value

In sentences such as the following, English allows no choice of form:

- (40) The five applicants arrive tomorrow.

Similarly in Hungarian predicate agreement with numeral phrases does not allow an option. The form, however, differs from that of English:

Hungarian (Edith Moravcsik, personal communication)

- (41) hat fiú érkez-ett
 six boy[SG] arrive-PST[3SG]
 'six boys have arrived'

The plural of *fiú* 'boy' is *fiúk*, and the plural of *érkezett* 'arrived' is *érkeztek*; neither would be used in (41).²³

In Russian, the situation is more complex. Let us take just one type: these Russian examples are both fully acceptable:

Russian

- (42) voš-I-o pjat' devušek
 come.in-PST-N.SG five[NOM] girl[PL.GEN]
 'five girls came in'
- (43) voš-I-i pjat' devušek
 come.in-PST-PL five[NOM] girl[PL.GEN]
 'five girls came in'

The essential point here is that, given the same controller, target, domain and feature specification of the controller, there remains a choice of agreement. Taking a set of the quantifiers, I counted all relevant examples in a corpus of texts from the nineteenth and twentieth centuries (details in Corbett 1983: 150–3) and found 235 relevant examples, of which 54% showed singular agreement as in (42) and 46% showed plural agreement as in (43). (We return to conditions on the choice

²³ Amharic combines the possibilities of English and Hungarian. In construction with a quantifier, a noun may be singular or plural. If the noun phrase is subject, the verb then agrees, being singular if the noun is singular, and plural if it is plural (Leslau 1995: 179–80).

in §1.4.5.) Therefore the situation found in English and Hungarian (no choice of feature value) is more canonical than that found in Russian (choice of value).

This criterion links to the last, but is distinct from it. While choices typically involve semantic agreement in one option, semantic agreement may or may not involve an agreement choice for a particular target. For instance, in the example

(44) this man and woman have travelled all day to meet you

The use of *have*, the result of number resolution, is an instance of semantic agreement, but is obligatory (at least for some speakers).

While many accounts ignore them, agreement choices are rampant. In §5.5, I investigate the factors which can give rise to them. As we shall see in chapter 7, however, while choices are frequent, the variation we find is far from random.

1.4.5 Conditions

Here the criterion is straightforward and intuitive:

C-20: no conditions > conditions

That is to say, in the canonical situation, when the controller, target, domain and features have been specified for a particular agreement construction, that constitutes a full specification. If we need in addition to specify a condition, that is less canonical. For example, we noted the agreement choice in examples (42) and (43) above. There is good evidence that controllers denoting animates in such constructions are more likely to take agreement forms with a greater degree of semantic justification (plural here) than are those referring to inanimates. Similarly, controllers which precede their targets are more likely to take agreement forms with a greater degree of semantic justification than are those which follow. Chapter 6 is devoted to conditions on agreement and so I can be brief here. We should note, however, that agreement conditions are particularly prevalent when agreement is non-canonical in some other way. In the Russian examples the condition interacts with an agreement choice, itself a non-canonical characteristic.

1.4.6 Three general principles

Three general principles were introduced earlier, and deserve brief discussion here. It is important to note that they never conflict; on the contrary, like all the criteria which they cover, they converge on the notion 'canonical agreement'. We have no need to rank them nor to specify what happens in situations of conflict, because the criteria are mutually compatible.²⁴

Principle I: Canonical agreement is redundant rather than informative

²⁴ Canonical is thus a more abstract notion than prototypical; canonical can be clearly defined, but in principle there need not be a real instance, only approximations, while prototypical implies real instances.

Several separate criteria (numbers 1, 2, 10, 17, 18 and 19, and secondarily number 16) converge on this principle. It may be that it is this principle which leads to canonical agreement being relatively rare among the world's languages.

As a partial restating of this principle, we might add that the greater the reliance on formal properties the more canonical the agreement. This view of it is best seen by imagining its opposite. If we had fully semantic agreement, then it would hardly exist as a distinct phenomenon, since all the forms could be predicted directly from semantics; the matching effect would arise from controller and target corresponding simply by virtue of having a common semantic source. It is in the converse cases, for example in agreement in gender in instances where the gender is not assigned by a semantic rule, that we most evidently require special rules of agreement.

Principle II: Canonical agreement is syntactically simple

This principle is that canonical agreement can be described in straightforward rules, while non-canonical instances typically involve an additional complication. It is a generalization of criteria numbers 3, 4, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20 and partially of number 6. It is reflected in criteria relating to each aspect of agreement (controller, target and so on).

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

Different criteria converge on this principle, namely numbers 5, 6, 7, 8 and 9. Note that they all relate to the target. There are different views as to which target types are legitimately considered to be a part of agreement, but no-one, I think, would exclude the type of targets with canonical inflectional morphology from an account of agreement. The criteria which fall under this principle have application beyond agreement, in that they are part of a typology of inflectional morphology, based on canonicity.

I wished to clarify some of the conceptual problems and misunderstandings that characterize this area. We have seen how different properties cluster, which makes it particularly important that we specify which properties are the basis for our analytical decisions. Seeing the gradient nature of many of the properties (as well as the ways in which they overlap), makes the question of 'drawing the line' between agreement and other phenomena appear secondary. It is more important to understand agreement and its related phenomena than to draw a precise line at which we might claim agreement 'stops' and some other phenomenon begins.

1.5 Scope and structure of the book

Given the importance of the topic and its diversity across languages the book could have run to many volumes. It could have been filled with details