98.–99. Alignment of Case Marking

Bernard Comrie

1. Main alignment types

These two maps are concerned with the ways in which core argument noun phrases are marked — by means of morphological case or adpositions — to indicate which particular core argument position they occupy. The first map is concerned with full noun phrases, the second with pronouns.

The core argument of a canonical, one-place intransitive predicate may be symbolized S. The two core arguments of a canonical, two-place transitive predicate may be symbolized as A and P, with A representing the more agent-like argument and P the more patient-like (Comrie 1978). (In another terminology, the symbol O is used rather than P; Dixon 1994.) In studying the alignment of case marking, we ask the question which of S, A, and P are coded identically and which are coded differently. Note that for the purposes of this chapter, only case marking is considered. Alignment of person marking in the verb is treated in chapter 100. Other manifestations of alignment are also possible, such as word order, but are not treated here.

In the **neutral** case marking system, all of S, A, and P are marked in the same way. This can be illustrated by Mandarin examples (1a-b), where neither the S of (1a) ('the person'), nor the A of (2b) ('Zhangsan'), nor the P of (2b) ('Lisi') receives any case marking.

(1)   Mandarin (Li and Thompson 1981: 20)
    a.   *rén   lái   le*
          person come CRS
          'The person has come.'
    b.   *zhāngsān   mà   lǐsì   le   ma*
          Zhangsan   scold Lisi   CRS Q

'Did Zhangsan scold Lisi?'

In the **nominative–accusative** (or simply: accusative) case marking system, S and A are marked in the same way, while P is marked differently. The form used to encode S and A is referred to as the *nominative*, the form used to encode P as the *accusative*, as illustrated in Latvian examples (2a-b).

(2)   Latvian (Mathiassen 1997: 181, 187)
    a. *Putn-s   lidoja.*
        bird-NOM   fly.PST.3
        'The bird was flying.'
    b. *Bērn-s   zīmē   sun-i.*
        child-NOM   draw.PRES.3 dog-ACC
        'The child is drawing a dog.'

Note that the definition of the nominative–accusative system says nothing about *how* the distinction between S/A and P is marked. In Latvian, both nominative and accusative have overt markers. However, it is also possible for just the accusative to have an overt marker, as in Hungarian, where the word for 'person' is *ember* in the nominative, but *ember-t* in the accusative. Much less frequently cross-linguistically, it is the nominative that has an overt marker and the accusative that lacks one, as in Harar Oromo (Cushitic, Afroasiatic; Ethiopia) examples (3a-b).

(3)   Harar Oromo (Owens 1985: 101, 251)
    a. *sárée-n adíi-n   ní   iyyi-t-i*
        dog-NOM white-NOM   FOC   bark-F-IMPF
        'The white dog is barking.'
    b. *haat-tíi   okkóttée   goot-t-i*
        mother-NOM   pot   make-F-IMPF
        'Mother is cooking (lit. making the pot).'

Since the "**marked nominative**" type illustrated by Harar Oromo is a topic of current typological and theoretical interest, it has been given a separate encoding in the maps, contrasting with the standard type where either just the accusative or both nominative and accusative are marked.

In the **ergative-absolutive** (or simply: ergative) system, S and P are encoded in the same way, and A is encoded differently, as in Hunzib (Nakh-Daghestanian; eastern Caucasus) examples (4a-b).

(4) Hunzib (van den Berg 1995: 122)
　　a. *kid　　y-ut'-ur*
　　　　girl　　CL2-sleep-PST
　　　　'The girl slept.'
　　b. *oždi-l　　kid　　hehe-r*
　　　　boy-ERG　　girl　　hit-PST
　　　　'The boy hit the girl.'

The case that encodes S and P is referred to as the *absolutive*, the case that encodes A as the *ergative*. (In an alternative terminology, the case that encodes S and P in the ergative-absolutive system is referred to as the *nominative*. This usage is not adopted here, to avoid confusion.) In Hunzib, the ergative case has an overt marker, *-l*, while the absolutive does not. However, it is also possible for both cases to have overt markers, as in Tukang Besi (Western Malayo-Polynesian; Indonesia; Donohue 1999: 51), where the preposition *na* marks the absolutive, the preposition *te* the ergative. The "marked absolutive" is exceedingly rare, having been so far attested only in one language, Nias (Western Malayo-Polynesian; Indonesia), where the absolutive is marked by modifying the initial segment of the ergative (Brown 2001).

In the **tripartite** system, all of S, A, and P are marked differently. This system is found for some noun phrases in Hindi, as illustrated in examples (5a-b).

(5) Hindi (Anvita Abbi, p.c.)

  a. *laRkaa   kal   aay–aa*
     boy   yesterday come.AOR-SG.M
     'The boy came yesterday.'

  b. *laRke   ne   laRkii   ko   dekh–aa*
     boy.OBL ERG girl   ACC   see-SG.M
     'The boy saw the girl.'

In (5a), the S has no overt marker. In (5b), the A has the ergative postposition *ne* (which requires the preceding noun to be in the oblique case), while the P has the accusative postposition *ko*. For only one language has it been claimed that all noun phrases have the tripartite system, namely Warrungu (Pama-Nyungan; Australia; Tasaku Tsunoda, p.c.).

There is one other logical possibility for grouping S, A, and P, namely for A and P to have the same form, while S has a distinct form. This possibility is exceedingly rare; it does not occur in our sample, but is attested in some Iranian languages of the Pamir region, though restricted to some pronouns (Payne 1979).

In all of the systems discussed so far, there has been consistent encoding of all instances of S in the same way. However, another possibility is for S to be split between more agent-like and more patient-like instances of S, which we may symbolize as S_a and S_p respectively. On the basis of semantic similarity, S_a then groups with A, while S_p groups with P. This system has come to be called the **active–inactive** (or simply: active) system, on the basis of terminology originally created by the Russian linguist Georgij A. Klimov, though other terms are also found, e.g. *agentive-patientive* or *stative-active*. The active form covers S_a and A, the inactive S_p and P. This system is rather widespread as a basis for person marking on verbs (see chapter 100), but it is also found occasionally with case marking, as in examples (6a–c) from Georgian.

(6)   Georgian (Harris 1981: 40)

    a. *vaxt'ang-i   ekim-i   iqo*
       Vakhtang-PAT doctor-PAT be.AOR.3SG
       'Vakhtang was a doctor.'

    b. *nino-m   daamtknara*
       Nino-AGT   yawn.AOR.3SG
       'Nino yawned.'

    c. *nino-m   ačvena   surat-eb-i*
       Nino-AGT show.AOR.3SG>3SG>3SG picture-PL-PAT
       gia-s
       Gia-DAT
       'Nino showed the pictures to Gia.'

In (6a), the more patient-like S of the copular verb stands in the inactive case, while in (6b) the more agent-like S of 'yawn' stands in the active case. The active case is also used for the A of (6c), the inactive case for its P. The active-inactive system is rare for case-marking, and is identified in only four languages of the sample for these chapters: Drehu (Oceanic; New Caledonia; Moyse-Faurie 1983: 147), Basque (Hualde and Ortiz de Urbina 2003: 364), Georgian, and Imonda (Border family; Papua New Guinea; Seiler 1985: 138-139). The available examples are consistent with the criteria distinguishing active from inactive clauses with respect to person marking on verbs, such as volitionality and dynamicity (Mithun 1991).

    Finally, the system "**none**" is restricted to pronouns and is used for languages where pronouns are not permitted in one or more of the positions S, A, and P; in the sample used for the chapters these languages are Wari' (Chapacura-Wanham; Brazil; Everett and Kern 1997: 122) and Wichita (Caddoan; Oklahoma; Rood 1976: 10) (in both of which personal pronouns are not found for any of S, A, and P) and Canela-Krahô (Macro-Ge; Brazil; Popjes and Popjes 1986: 175), in which personal pronouns are found in S and A position, but not in P position.

2. Map 98: Alignment of case marking of full noun phrases

@	1.	Neutral	98
@	2.	Nominative-accusative (standard)	46
@	3.	Nominative-accusative (marked nominative)	6
@	4.	Ergative-absolutive	32
@	5.	Tripartite	4
@	6.	Active-inactive	4
		total	190

In deciding which type to assign a particular language to with regard to its case marking alignment, a number of problems arise, and as far as possible an attempt has been made to find consistent, or even better: principled solutions to these problems, although a number of difficult cases remain. In this section, problems relating to full noun phrases are treated. Most of these same problems also carry over to pronouns, but pronouns introduce a further set of difficulties, which form the topic of §3.

First, there is a general problem concerning the dividing line between the ergative (including the tripartite) and active systems. In a number of languages that have a basically ergative system, a small number of intransitive verbs, or a small semantic range of intransitive verbs, require their S to be in the case identified as ergative. Such languages thus stand between a pure ergative system (where all intransitive verbs take their S in the absolutive) and a pure active system (where intransitive verbs are divided into two substantial sub-sets, one taking S in the active, the other taking S in the inactive). The policy adopted here is that for a language to be considered of the active type, there must indeed be two substantial sets of intransitive verbs differing in the case marking of their S. A small number of exceptional intransitive verbs, as in Hindi (McGregor 1977: 73–

74), or a small semantic area of intransitive verbs (as with onomatopoeic verbs in Hunzib; van den Berg 1995: 124), will not be taken into account, so that Hindi is considered tripartite, Hunzib ergative. Sometimes, there are dialect differences within a language; thus, both Basque and Georgian dialects differ in how many intransitive verbs take active subjects.

But the main recurrent difficulty is that in many languages, different kinds of full noun phrases partake of different case marking patterns. For instance, in Spanish the accusative marker, the preposition *a*, is found (roughly) only with specific, animate noun phrases, so that strictly speaking a noun phrase like the male proper name *Juan* has a nominative-accusative case marking system, while the inanimate noun phase *el libro* 'the book' has a neutral case marking system, as illustrated partially in (7).

(7)   Spanish
        a. *María   vio   a   Juan.*
            Mary   see.AOR.3SG   ACC John
            'Mary saw John.'
        b. *María   vio   el   libro.*
            Mary   see.AOR.3SG   the book
            'Mary saw the book.'

(Instances where the P sometimes takes case marking, sometimes not, have come to be called *differential object marking*.) In some languages, a case marker is used primarily to avoid ambiguity, so that in Lower Grand Valley Dani (Trans-New Guinea; Papua, Indonesia; Bromley 1981: 84–85), for instance, the ergative marker is used in this way. In yet other languages, case markers may be described as optional, without any detailed discussion of the conditions under which the marker does or does not occur, e.g. the accusative marker in Burmese or the ergative marker in Araona (Tacanan; Bolivia; Pitman 1980: 14). More complex patterns may arise, e.g. in Gooniyandi (Bunaban;

Australia) the ergative marker is optional with animate nouns but obligatory with inanimate nouns (McGregor 1990: 319–320); in Hindi, the accusative marker is used for Ps that are higher in definiteness and especially animacy, while the ergative marker is obligatory for A. The policy that has been followed in assigning such languages to types has been to maximize the occurrence of overt case marking. Thus, if a language has an optional accusative case marker, or one that occurs only under certain specified circumstances, then this has been given priority and taken as criterial. This policy decision needs to be taken into account consistently in interpreting the maps. (For details on how decisions were taken for individual languages, reference should be made to the electronic version of this atlas.) Thus, Spanish and Burmese come out as accusative, Araona and Gooniyandi as ergative, and Hindi as tripartite.

In a number of languages, the case marking system is different in different tense-aspect-moods (TAMs). In Georgian, there are three sets of TAMs with respect to the fine details of case marking, grouping into two sets with respect to the factors relevant to present concerns. In the Aorist and Perfect series, case marking is on an active-inactive basis. However, in the Present series, it is on a nominative-accusative basis. In Hindi, the tripartite system is found in the Perfective aspect, while the Imperfective has a nominative-accusative system (with, as noted above, the further complication that accusative marking depends on animacy and definiteness). A particularly complex system is found in Drehu: in the Non-Past, the "agent" marker is obligatory for A and optional for S_a, i.e. the system is, in terms of our criteria, active-inactive; in the Past, this marker is used for all S and A (with some exceptions for inanimates), i.e. the system is in our terms nominative-accusative. In such cases, our general policy has been to maximize the occurrence of otherwise cross-linguistically rare types. Thus, I assign Georgian to the active type, effectively giving preference to the Aorist/Perfect over the Present; Hindi to the tripartite system,

giving preference to the Perfective over the Imperfective; and Drehu to the active type, giving preference to the Non-Past.

Finally, some languages have two distinct, voice-like constructions with different case marking alignment systems, where moreover there might be controversy as to which system should be taken as most basic. In a language with a voice system where one voice is clearly more basic than the other, as with basic Active and non-basic Passive in English, I take the basic voice. But some other languages, especially some Austronesian languages, are less clear. For Tukang Besi, Donohue (1999: 53) argues that the transitive construction with ergative-absolutive case marking is more basic, relative to an alternative construction with nominative-accusative case marking. (A similar problem of identifying the basic voice arises in Karo Batak (Western Malayo-Polynesian; Indonesia; Woollams 1996), but fortunately both voices have neutral case marking.) In some languages, the decisive factor is word order. For Rapanui (Oceanic; Easter Island), Du Feu (1996: 67-68) argues that the basic word order is VAP; in this order, neither A nor P is overtly case marked, and the system is neutral. However, in the alternative word orders VPA and PVA, the A requires the ergative preposition *e*, and case marking is thus ergative. In Paumarí (Arauan; Brazil; Chapman and Derbyshire 1991: 164-166), the usual word order in transitive clauses is AVP, and the case marking is ergative-absolutive; however, under alternative word orders PVA and APV, the case marking is nominative-accusative. (An added complication in Paumarí is that object pronouns can only be preverbal, so that with pronouns only the nominative-accusative system is possible.) In such cases of voice-like alternations, including those described as conditioned primarily by word order, I have followed the author of the source in deciding which alternant is basic, and have used the case marking system of that alternant. (In the case of Tagalog (Western Malayo-Polynesian; Philippines), I took as decisive

constructions in which neither S nor A nor P is in "focus", which leads to a neutral system.)

3.　　Map 99: Alignment of case marking of pronouns

@	1.	Neutral		79
@	2.	Nominative–accusative (standard)		61
@	3.	Nominative–accusative (marked nominative)		3
@	4.	Ergative–absolutive		20
@	5.	Tripartite		3
@	6.	Active–inactive		3
@	7.	None		3
			total	172

　　　　Pronouns merit a separate map from full noun phrases because in many languages pronouns have a different case marking system from full noun phrases — 23 languages in the sample used for these chapters (25 if the marked nominative system is considered distinct from the rest of nominative–accusative). In English, for instance, while full noun phrases have a neutral case marking system, pronouns have a nominative–accusative system, e.g. with *we* for S and A but *us* for P. Some of the problems discussed in §2 apply also in the case of personal pronouns, and in such instances similar solutions are applied, with one additional criterion, namely the assignment of some weight to differentiating pronouns from full noun phrases. Thus, in English some pronouns have a nominative–accusative system, while others (in particular, second person *you*) do not; the nominative–accusative system is selected here in part because it differs from the case marking system for full noun phrases. Sometimes the problems that occur with full noun phrases occur with somewhat less force in the case of pronouns; for instance, several languages where full noun phrase Ps sometimes take an accusative marker and

sometimes not require pronouns to take this marker (e.g. Persian), so that there is no hesitation in assigning them to the nominative-accusative type with respect to case marking of pronouns. However, pronouns bring with them a host of other problems that can lead to uncertainties in assigning languages to case marking alignment types with respect to their pronominal systems.

These problems center on the question of what constitutes a pronoun. In some languages, such as English, pronouns have essentially the same distribution as other noun phrases, so that there is direct comparability between pronouns and full noun phrases with respect to case marking alignment. However, this is overall a clear minority pattern among the world's languages; in the vast majority of the world's languages it is, for instance, not necessary to have a pronoun in subject position (see chapter 101).

For some of the languages in the sample, we have clear statements in the sources consulted that personal pronouns, at least in some of the S, A, and P positions of the clause, are simply impossible: Canela-Krahô, Wari', Wichita (see §1). For many other languages, the sources note that pronouns are rarely used in the language in question, but nonetheless the sources are reasonably clear that if pronouns are used (for instance, for contrast), then they have such-and-such a case marking system. In yet other languages, the sources are not sufficiently explicit to enable the reader to reach a firm conclusion, with the result that several languages included in the map for full noun phrases are not included in the map for pronouns. In many of the languages covered by the present paragraph, the function of expressing the person-number of the S, A, or P of a clause is covered by person-number marking on the verb, for instance in Abkhaz (Northwest Caucasian; Abkhazia, Georgia). And at least for some such languages, it has been argued that the "real" S, A, and P are not the pronouns, but rather the pronominal affixes on the verb (which fall under the

topic of chapter 100 rather than of the present chapters). However, there are also languages like Japanese which tend to avoid using pronouns but nonetheless have no encoding of the corresponding person-number information in the verb.

The real problem is caused by elements that encode such pronominal features as person-number but whose status on the scale from pronoun (as a kind of noun phrase) to pronominal marker (as a bound morpheme attached to the verb or other predicate) is unclear. In general, I have here followed the analyses of the sources, mainly because in general I have no reason to doubt these sources, although one must bear in mind that it can take a fairly sophisticated (and potentially controversial) analysis to really tease the various possibilities apart. In some instances, the analysis adopted here may be pushing against the boundaries of plausibility, for instance in treating the French clitic pronouns as relevant instances of pronouns (and thus concluding that French has a nominative-accusative case marking system for pronouns, on the basis of clitic oppositions like nominative *je* versus accusative *me*, although the disjunctive pronoun *moi* shows no case distinction). In some instances, then, more systematic research into the nature of pronouns may require some redrawing of boundaries.

4.   Geographical distribution

Some of the systems are so rare that it is questionable whether anything reliable can be said about their geographical distribution. Thus, the three instances of languages that do not allow pronouns in all of S, A, and P positions are all from the Americas, but this is surely coincidence; another such language (not in the sample) is the African language Mbay (Keegan 1997: 62-63). The three instances of the active system are more scattered, but again it is not clear that anything significant attaches to the fact that one is found in the Pyrenees (Basque),

one in New Guinea (Imonda), and one in Oceania (Drehu). The tripartite system emerges primarily from the intersection of nominative-accusative and ergative-absolutive systems, and is represented by only three languages in the sample for this chapter (Hindi, Nez Perce, and Semelai).

The marked nominative has a restricted though more interesting distribution. It is found first of all in Africa, both in Afroasiatic languages (Middle Atlas Berber and Harar Oromo in the sample), as well as in some Nilo-Saharan languages of East Africa (Murle in the sample) — this could well be a combined genealogical-areal grouping. In a quite separate part of the world, it is found in the Yuman languages of southern California, western Arizona, and northern Baja California. Two other languages in the sample are identified as marked nominative, although it is not clear that they instantiate exactly the same phenomenon. Igbo has complex tonal interactions between subject and verb that give rise, in some TAMs, to at least the appearance of a marked subject, although the traditional analysis would speak rather of floating tones marking TAM that happen to dock onto the subject. In Aymara, the accusative is formed by deleting the final vowel of the nominative, so that although the nominative has more phonetic material than the accusative, the direction of derivation seems rather to be from nominative to accusative (the vowel would not be predictable under the reverse derivation).

The neutral system is widespread across the world. While it might be expected in areas that otherwise have little morphology, such as Southeast Asia and West Africa, it is in fact also found in languages that have complex inflectional morphologies, but where such morphology is largely confined to the verb, as for instance in Bantu languages and several languages of the Americas. The accusative system is also widespread on a global basis. Although ergative case marking is also quite widespread, it is almost completely lacking from Africa and is rare in Europe; hotbeds of ergativity include

Australia and the Caucasus, to a somewhat lesser extent parts of the Americas, New Guinea, South Asia, and the Austronesian family.

5. Theoretical issues

Much of the theoretical interest surrounding case marking alignment has been concerned with correlations with other parameters, and thus features only indirectly in the present chapters. Comparison with chapter 100 enables a comparison between the incidence of alignment patterns with noun phrases and with person markers on verbs. The scope of the *World Atlas of Language Structures* does not permit comparison with other syntactic properties, including in particular the behavioral properties of S, A, and P.

Another area of theoretical interest that goes beyond the scope of this chapter is the set of factors that determine the use of a particular case marking pattern where the language in question allows alternatives (see, for instance, Silverstein 1976); this is because of the decision, noted in §2, to abstract away from such differences in assigning languages unequivocally to one type. Relevant generalizations would be the use of the animacy and definiteness hierarchies to constrain possible distributions of accusative and ergative case marking, with accusative case marking normally only used on a given noun phrase type if all noun phrase types higher on the animacy and/or definiteness hierarchies have accusative case marking; and conversely, with ergative case only used if all noun phrase types lower on these hierarchies have ergative case marking.

However, this claim can be tested on the basis of the present materials with regard to the relation between full noun phrases and personal pronouns, with personal pronouns, especially of the first and second persons, usually being claimed to be higher on the animacy hierarchy than full noun phrases, and thus more likely to have accusative case marking and less

likely to have ergative case marking. This would predict that, if languages have a difference between the case marking alignment of full noun phrases and pronouns, then the possible combinations should be (respectively) neutral (full NPs) + accusative (pronouns), ergative + accusative, and ergative + neutral, but not accusative + neutral, accusative + ergative, and neutral + ergative. And indeed this hypothesis is largely borne out by the languages forming the sample for these chapters. There are 9 languages with the combination neutral + accusative, 5 with the combination ergative + accusative (4 of them in Australia; in the fifth, Paumarí, as noted in §2, the difference correlates with different word order possibilities), and 6 with the pattern ergative + neutral (2 of them in the Caucasus and 2 of them Eskimo languages). There are, however, some exceptions: 2 languages with the combination accusative + neutral (though one — Middle Atlas Berber — has the marked nominative subtype for nouns, i.e. a somewhat aberrant variant of nominative–accusative case marking); no languages with the combination accusative + ergative; and only 1 language with the combination neutral + ergative — this last language is Chamorro, which has a complex interaction of case marking of pronouns and voice–like phenomena. The Marathi combination tripartite + accusative is like ergative + accusative and ergative + neutral in having a distinctive ergative case for nouns but not for pronouns.