

DEFINING SURFACE CASE

This paper is part of a research project in progress dealing with morphological case. We do not intend to deal here with the semantic properties typical of case, nor with predicate roles and valencies, i.e. what are often referred to as "deep cases" in contemporary linguistics, but rather will restrict the very term "case" to denote overt markers on nouns as we know them from the classical Indo-European languages and most modern Slavonic languages.

A case marker will here be considered as an overt morpheme 1) as opposed to other means of signalling grammatical relationships typical of nouns, such as word order or prepositions. More specifically we shall define a case marker as a discrete language sign, a bound morpheme that is obligatorily attached to a noun (substantive) stem via inflexion, most typically as a suffix. Optional semantic markers attached to nouns as in Tibetan may represent a stage of development towards a case system, but until they become obligatory, they cannot be analyzed as elements of a case system. The definition implies that if more than one noun occurs coordinated within a single NP, then every noun stem will be separately marked for case, if the marker is a true case morpheme.

Normally, a case marker will show agreement or concord with other forms occurring within the noun phrase. We do not claim agreement to be a necessary feature of case languages, but it is no doubt an important criterion. As will be seen in what follows it might be questioned whether languages such as Turkish, Mongolian and Tamil, all traditionally taken to be case languages without agreement, really ought to be regarded as proper case languages.

Though case by way of agreement indirectly may be assigned to a noun phrase, it is important to observe that the case marker itself is attached to the head noun. This is actually one of the phenomena differentiating proper case languages from caseless ones. In a caseless language a functional marker, semantically corresponding to the case morpheme of a case language, is typically attached to the noun phrase, but not to the head noun of that phrase, but rather either as a prefix (preposition) to the whole phrase or, occasionally, as a suffix (ending, postposition) attached to the last word of the noun phrase. Accordingly, the so-called *s*-genitive of English and Swedish is not a case marker. In both languages, it is true, the ending may be added to the head noun followed by a prepositional attribute in a very formal style and thus be treated as a true case marker: *the King's of Sweden court*, *Konungens av Sverige hov*. This type of construction is, however, nowadays practically obsolete and most speakers use the so-called group-genitive construction: *the King of Sweden's court*, *kungen av Sveriges hov*, where the marker *s* functions as a postposition equivalent to an ordinary preposition determining the whole noun phrase. It should be observed that Danish does not even in the most formal style allow the "genitive" *s* to be treated as a case marker: *Dronningen af Danmarks snue* (**Dronningens af Danmark*) 'The queen of Denmark's cold', *Kongen af Danmarks bolsjer* (**Kongens af Danmark*) 'The King of Denmark's Cough-Drops'. The alternative where a language that otherwise exclusively employs prepositions for signalling grammatical relations, nevertheless develops postpositions with a similar syntactic function is not very common. We know, however, of at least one language where an equivalent situation is present, viz. Persian. This language has a postposition *râ* (colloquially *ro* or *re*), which is traditionally held to signal the accusative (cf. Lazard 1957, 70, 175ff.; Windfuhr 1979, 47ff.): *fars dozd-râ gereft* 'fear seized the thief'. This *râ* is actually added to the last element in the noun phrase, be it a single noun, a noun with adjectives, or even a phrase with a verbal noun or an infinitive: *xâne-je hasan-râ mišenâsam* 'I know Hassan's house', the so-called *ezâfe -je* signalling that *hasan* 'Hassan' is an attribute to *xâne* 'house' (the object).

The assertion that case markers are attached to noun stems and that they do not determine noun phrases is a stipulative one. The purpose of this working definition is two-fold. On one hand we intend to develop a theory of morphological case, the outlines of which have been presented in Bily & Pettersson (1983; 1984), and explored to a certain extent in Bily & Pettersson (forthcoming), that is not only a case theory analogous to those of Hjelmslev

(1935) and Jakobson (1936; 1958). We intend to develop our theory in such a way that it explains what case intrinsically is. Secondly, we would like our theory to be able to form a firm basis for at least part of language typology. There are intrinsic structural differences between case and caseless or, as we shall prefer to call them in what follows, filter languages which are not restricted to the existence vs. non-existence of case morphemes as such. There are also a lot of differences in syntax and semantics between the two classes of languages which require exploration and explanation. With these considerations in mind it is obvious that we need a safe, mechanical procedure for distinguishing case markers from other types of morphemes. It will be seen below that our working definition has more advantages in addition to merely being a handy tool for classification.

Quite generally, the difference between the two classes of languages could be characterized as depending upon two different structuring principles. Whereas grammatical relations in filter languages can be depicted as a set of structural schemata with slots for different functions (as in Diderichsen 1946; for a discussion see Pettersson 1983), the corresponding functions in a case language form an abstract system of entities and relationships which exist, as it were, independently of relations present in actual speech. Thus in a filter language the concept of, for example, direct object does not exist outside of appropriate syntactic structures; it is mainly a matter of word order and intonational patterning. The accusative form of a noun of a case language belongs, on the other hand, to the lexicon as other morphemes and words of the language in question. Consequently, it exists independently of syntactic relationships. With this in mind it becomes obvious why we cannot accept Hjelmslev's description of a modern language such as English as a case language (Hjelmslev 1935, 118ff.) with, for example, a dative case marked by the word order or by certain prepositions. The stipulation that a case marker must determine a noun stem and not a noun phrase is helpful in clarifying such points. Word order is no morpheme, though it can fulfill functions similar to those of case morphemes, and prepositions are not attached to noun stems but determine noun phrases.

Our definition of case is very strong and would, at first glance, seem to raise problems. Consider for example modern Standard German. This language is a case language according to our criteria but, *nota bene!*, only due to the fact that it possesses a very restricted set of markers on certain noun stems, most consistently in the masculine and neutre genitive singular and the

dative plural. Otherwise case marking in German is overtly signalled only on determiners (articles or adjectives). The German articles, however, are not joined to noun stems but are prefixed to noun phrases and should therefore according to our definition not qualify as case markers! Rather than being disturbed by such complications, we on the contrary welcome them. Since our theory partly aims at formulating a working language typology, where the two classes both exhibit distinctive and unrelated structural properties, we should expect mixed forms to occur. In fact, German shows features of both classes: it is simultaneously a case language due to its ability to signal case on noun stems and a filter language in its having developed firm structural schemata. Furthermore, there are German dialects that have lost their ability to assign case to the noun stem. Such dialects have simultaneously lost case marking on the determiners as well, which is what one would predict given our definition of case. In fact, Standard Dutch could be taken as representative of such dialects. Here the old German plural has been generalized to a common plural marker, while the genitive singular has been retained but reinterpreted as a postposition as in English and Swedish or, more commonly, been replaced by a prepositional phrase or a construction with a special possessive marker (originally a possessive pronoun). Thus 'my sister's book' is in Dutch either *mijn zusters boek* (postposition), *het boek van mijn zuster* ('the book of [from] my sister'), or, colloquially, *mijn zuster d'r boek* (lit. 'my sister her book'). The interesting point is that the articles have lost every trace of case marking: the definite article is *de* for masculine, feminine and plural nouns, and *het* for the neutre singular, while the indefinite article is *een* (optionally 'n) throughout. Thus German does not contradict our definition. Its articles show case due to agreement with the head noun. If the German head noun were to lose its ability of signalling case at some future date, the secondary case marking of the noun phrase could presumably not be upheld unless by force from prescriptive grammar.

Although German does not cause any serious problem for our definition, Roumanian ²⁾ would seem to come much closer to constituting a clearcut counter-example. Roumanian is traditionally held to possess two separate cases in the nominal system, the nominative-accusative (NA) and the genitive-dative (GD). When indefinite, the noun has no case marking, but in the definite form, case is overtly signalled according to the following paradigms for the nouns *lup* 'wolf' and *casă* 'house', representing masculine and feminine noun declensions respectively:

	Indefinite		Definite		
NA	<i>lup</i>	<i>casă</i>	<i>lupul</i>	<i>casa</i>	Sg
	<i>lupi</i>	<i>case</i>	<i>lupii</i>	<i>casele</i>	Pl
GD	<i>lup</i>	<i>casă</i>	<i>lupului</i>	<i>casei</i>	Sg
	<i>lupi</i>	<i>case</i>	<i>lupilor</i>	<i>caselor</i>	Pl

So far Roumanian would seem to constitute a real case language, the definite ending being a suffix attached to the noun stem. In reality, however, the article determines the noun phrase just as in English and German. Normally an adjective attribute follows its head noun: *limba română* 'the Roumanian language', *lupul mare* 'the big wolf'. Though much more restrictedly than in French, some adjectives may be used attributively preceding the noun. Thus 'the great poet' may be in Roumanian *poetul mare*, but the more usual expression is actually *marele poet*, where *-le* is the article. Likewise the article will be joined to the adjective in the GD case: *casa marelui poet* 'the great poet's house'. This demonstrates that Roumanian distributes its definite articles in exactly the same way as Bulgarian, viz. by adding the ending to the first word of the noun phrase and thus letting the article determine the noun phrase and not the noun contrary to the normal situation in Nordic languages; cf. Bulgarian *ezik* 'language', *ezikăt* 'the language' but *bălgarskiyat ezik* 'the Bulgarian language' as opposed to *bălgarski ezik* 'id. (indef.)'. Roumanian has furthermore an indefinite article, also taking the case marker: NA *un lup* 'a wolf', GD *unui lup*. Consequently, provided our working definition of case languages is correct, Roumanian is not a genuine case language. There is, however, one exception to the general rule that Roumanian exclusively signals case on the noun phrase and not on the noun. Feminine nouns preceded by a declined indefinite article have a special GD form: *prețul unei case* (**prețul unei casă*) 'the price of a house'. Moreover, an adjective attribute of such a GD form will show concord with its head; 'a big house' is *o casă mare*, whereas 'the price of a big house' is *prețul unei case mari*, where *mari* is the singular GD form of *mare* 'big' (it should be noted that these particular forms are identical to the indefinite plural form of the words in question). Though Roumanian could be accepted as a case language due to this peculiarity, we are not entirely happy with the situation. In other respects Roumanian does not clearly exhibit certain properties that we consider typical of case languages; its prepositions, for example, normally do not govern the GD case but rather the NA case.

Consider, furthermore, that the Roumanian GD forms of the definite articles are morphologically identical to the GD forms of the third person personal/possessive pronouns *lui, ei, lor*, 'him/his, her, them/their' (Roumanian has a well developed pronominal case system). This fact then allows us to interpret the declined Roumanian definite articles as possessive markers corresponding to the possessive markers of colloquial Dutch, which in turn means that we can claim that the Roumanian articles are part of the pronominal system. ³⁾ The Roumanian expression for 'my sister's book', *sorei mea carte*, can therefore be interpreted as rendering a structure like 'my sister her book', which, word for word, corresponds to the Dutch *mijn zuster d'r boek*. What then remains of the Roumanian nominal case marking is the specific case form allotted to feminine nouns following the declined indefinite article. However, since the appearance of this GD form is triggered only by the GD form of the indefinite article we have good reason to question even this form as a proper case marker. In accordance with Mel'čuk (1981) we will claim that even if agreement or concord signals morphological dependence, it does not necessarily express the direction of the dependence in question, be it syntactic or semantic. We may, therefore, turn the argument upside down and say that the feminine head noun and any adjective determiners agree morphologically with the oblique form of the indefinite article. Thus not even this special instance of apparent case marking in Roumanian conflicts with our analysis of Roumanian. It is not a case language.

On the other hand, if our definition of genuine case languages had been invalidated by the occurrence of an undisputed case language showing all other properties of case languages except for the attachment of the case marker to the head noun, we would of course have to accept Roumanian as a case language. The GD agreement with the oblique indefinite article may then be interpreted in the usual way as an agreement of determiners with the head noun or as a concord phenomenon as will be explicated below. However, the preceding discussion makes it obvious that we would be far from happy with such a solution.

There is another peculiarity about Roumanian that has also been put forward as an argument in claiming its status as a full-fledged case language. Under certain circumstances, mainly when the noun denotes a person but also in some other cases, a particle *pe*, which is in fact a preposition, is obligatorily placed in front of the direct object: *am întâlnit pe Ion* 'I have met John', *am văzut pe vecin* 'I have seen the neighbour'. This "accusative" marker *pe* is used even before a personal pronoun

already inflected for the accusative: *ma văzuț pe mine* 'he has seen me', where *mine* 'me' is a specific accusative form of the first person singular personal pronoun distinct from the nominative *eu* and the dative *mie*. Since the particle determines noun phrases - *cunoaște pe un profesor* 'he knows a (certain) professor' - it does not qualify as a case marker according to our definition. In fact, though traditional grammar refers to the particle as "the accusative object marker *pe*", experts on Roumanian, e.g. Lombard (1973, 86), are not very apt to recognize it as a proper case morpheme.

The reason why we are including a discussion of the Roumanian *pe*, however, is the fact that Hebrew possesses a corresponding syntactic device, a direct object marker *et*, which much more categorically has been claimed to constitute a proper case marker. The Hebrew marker is required before any definite direct object, no matter what kind of entity the noun denotes: *dan ahav et rina* 'Dan loved Rina', *ani makar et hajéled* 'I know the boy', *rina bišla et hamarak* 'Rina cooked the soup', but *rina bišla marak* 'Rina cooked soup' (indefinite and consequently with no *et*). Moreover, the use of the Hebrew *et* is restricted to mark direct objects only. It cannot be used as an adverbial preposition as the Roumanian *pe*: *pe o masă* 'on a table'. Though in many respects exhibiting properties traditionally identified with the functioning of ordinary prepositions, the Hebrew *et* and, in fact, a set of other Hebrew prepositions as well, viz. *le*, *be*, *al*, *šel*, and *mi* (glossed as 'to', 'in, at', 'on, by', 'of', and 'from', respectively), behave in a way peculiarly similar to the case markers of e.g. Russian. Berman (1978, 119ff.), though discussing Hebrew case in Fillmorean terms and therefore a priori denying the relevance of surface case for grammars of natural languages, actually gives reasonable syntactic evidence for interpreting at least the above listed Hebrew prepositions as true case markers. Unfortunately, our own definition is not *prima facie* helpful with respect to Hebrew, since Hebrew is a language where the difference between nouns proper and noun phrases is not quite clear. For example Hebrew has a definite article *ha*, in fact a prefix, which is distributed as in Greek, i.e. attached to and thus determining each constituent within the noun phrase: *iša jafa is* 'a beautiful woman', while 'the beautiful woman' is *haiša hajafa*; cf. Greek *he guné he kalé* 'id.'. An adjective shows concord in gender and number with its head noun, but, again, the governing role of the head noun is signalled only with respect to gender and only when the noun has an atypical phonological form with respect to its morphological patterning, i.e. we find in Hebrew the type of mixed morphological paradigms that are usually exemplified with the

Latin *agricola bonus* 'a good farmer'. Berman asserts that the Hebrew prepositions determine noun phrases in spite of the fact that the initial element of a noun phrase will actually be its head noun. The reason why she nevertheless insists that the Hebrew particles in question are true prepositions and not case markers in our sense is that they do not affect modifiers of the noun such as adjectives or determiners, while other grammatical items such as number, gender, and species do. Since we, on the one hand, do not regard agreement as a necessary requirement for case marking but, on the other hand, are not quite happy with allotting Hebrew the status of a case language either, we feel that the agreement argument should be examined in more detail.

As a matter of fact, the problem we are faced with as far as Hebrew is concerned is the same one that is encountered with respect to the relational particles of Japanese. Japanese particles such as *wa*, *ga*, *ni* and so forth are placed after a noun phrase in such a way that it is impossible to tell whether the marker should be interpreted as determining the noun phrase as a whole or merely is attached to the head noun. The reason for our dilemma is obvious: neither Hebrew nor Japanese distinguish the noun phrase and the noun proper as different levels of sentence structure. This situation is very much the same in these languages as it is in the classical Indo-European languages Sanskrit, Greek, and Latin. The fact that the linguist is able to apply an analysis in terms of NPs and Ns to nominal constructions of these languages, of course, in no way means that the analysis also corresponds to the factual structure of the languages in question. On the contrary, there is in our opinion no indication that would favour such an analysis, since the same set of endings is used to mark all constituents within the noun phrase. As for Latin, Greek, and Hebrew the only exception seems to be gender in mixed paradigms. From such paradigms, however, it does not follow that concord should be interpreted in terms of different constituent levels. Rather, the situation should be described as follows. Whereas number and/or case and/or species in these languages are overtly expressed on all concatenated elements constituting a noun phrase, gender is overtly and obligatorily expressed only on such nouns that have no inherent gender. In other words, the languages in question possess two kinds of nouns, those with inherent gender (*nomina substantiva* in terms of medieval grammar) and those without inherent gender (*nomina adiectiva*), i.e. words which have been traditionally referred to as "adjectives". Consequently, as far as Latin noun phrases such as *populus Romanus* 'the Roman people', *puer parvulus* 'a little boy', *agricola bonus* 'a good farmer' are concerned, each phrase consists of a pair of conca-

tenated nouns, differing with regard to inherent gender but otherwise morphologically on the same foot. That is to say, the masculine ending *-us*, though a very frequent indication of masculine gender in *nomina substantiva*, is a proper gender marker only for nouns without inherent gender. Whether we call nouns of the latter kind adjectives or not is merely a matter of terminology. The important point is that, from a morphological and partly from a syntactic point of view, they are real nouns. It is for this reason that we have used and henceforth consistently will use the term "concord" to apply to such an equal correspondence in form between two nouns (or two words of the same word class) concatenated within the same phrase and denoting the same entity in the outer world, whereas the term "agreement" has been and shall be understood as referring to a correspondence in form such as it is known from modern Germanic and Slavonic languages. These languages possess special morphological means of signalling syntactic subordination within the noun phrase. Agreement in this sense simply does not exist in the classical languages. Thus the stem *bon-* in *agricola bonus* takes the masculine form not because it agrees with the noun *agricola*, which has inherent masculine gender, but because it refers to, denotes a male being. The same thing will be true of the categories number and case: any noun within the noun phrase will take just the form appropriate to the number and the specific roles of the entities referred to in the outer world. As can be seen the concord in case within the noun phrase is a necessary consequence of our definition of case markers as linguistic signs attached to noun stems but not to noun phrases.

We are now in the position of being able to solve the problem with the Hebrew relational particles. Hebrew has, just as Latin, two classes of nouns, those with inherent gender and those without. Both types are obligatorily marked for number but only the latter demands an overt marker for gender. For this reason we get mixed paradigms in Hebrew as well. The noun *sulxan* 'table' has inherent masculine gender but forms its plural with the "feminine" ending *-ot*; *gadol* 'big', on the other hand, has no inherent gender and, consequently, when used to refer to objects such as tables, recognized by the Hebrew linguistic society as male things, it must take the masculine ending proper *-im*. 'Big tables' is therefore in Hebrew *sulxanot gadolim*. Still, both words in the phrase are true nouns. Consequently, if the relational particles were true case markers they should have been attached to both words. For this reason and no other we can truly state that Hebrew is not a case language.

This assertion probably also holds true of Japanese and a host

of other languages which in the literature are held to possess morphological case. Often references to such "case" particles or endings are made quite loosely and do not seem to bear any particular theoretical salience. However, for two languages at least, viz. Turkish and Mongolian, tradition very stubbornly insists in claiming their status as full-fledged case languages. Turkish, for example, is supposed to possess six separate cases, viz. the absolutive or nominative, the definite accusative, the genitive, the dative, the locative, and the ablative. Some grammarians include in addition an abessive, a relative, and a comitative, whereas still others do not recognize the locative and the ablative as real cases (cf. Peters 1947, 27). No matter how many cases there are in Turkish, they are all expressed by endings attached to the final element of a nominal group with one single exception - the absolutive, which has no ending at all. Now, just as in Latin and Hebrew there is no morphological difference between an adjective and a noun. An adjective may therefore take the plural morpheme, any case ending or personal suffix just as an ordinary noun. It can also have the indefinite article *bir* before it. In contrast to Latin and Hebrew, however, Turkish, from a morphological point of view, consistently treats the noun phrase and not the noun or even the head noun as the finite immediate constituent. Therefore, the plural will never be marked more than once per nominal group: *bir kitap* is 'a book' and *kitaplar* is 'books', but 'three books' is *üç kitap*, *üç* 'three' being inherently plural. The same thing applies to all other grammatical morphemes, case markers included. The phenomenon is known as the principle of suspended affixation, a formulation that could be taken to imply that in Turkish, all affixes are present at some deeper level but are deleted during derivation. It should be observed that the suspension also applies to coordinated nouns as in *tebrik ve teşekkürlerimi sunarım* 'I offer my congratulations and thanks (lit. congratulation and thank-s-my-ACC offer-I)', where the grammatical morphemes *ler-im-i* (plural - possessive - accusative) are associated with both *tebrik* 'congratulation' and *teşekkür* 'thank' (Lewis 1975, 41). It seems to us just as natural to say that the morphemes in question are attached to the nominal group as an indivisible whole. If so, we could claim that Turkish is not a case language. Given our stipulative definition of case markers the Turkish case endings are of course nothing but postpositions equivalent to prepositions of e.g. English. On the other hand, it must be left to the specialist of Turkish to decide whether the principle of suspended affixation is really a feature of grammar and not merely a way of expression. Needless to say, we do not deny the possibility that the suspension idea may be the right one in handling the issue. In that case, Mongolian and Tamil would also be classified

as case languages, since they exhibit exactly the same pattern with respect to the noun phrase.

Originally we wanted to include Hungarian in the class of languages discussed above, Hungarian also being agglutinative and subject to the suspension principle just as Turkish. Having read Borin (1984), however, we have become a little more hesitant to do so. Borin, using our casehood definition as a point of departure, claims that Hungarian really is a case language. He comes to this conclusion by pointing to the fact that the two demonstrative pronouns *ez* 'this' and *az* 'that' obligatorily show concord with their head. In the sentence *Ekkor hallottam először ezt a nevet* 'Then I heard this name for the first time' (Borin's example), *ezt* is the accusative of *ez* and *nevet* is the accusative of *név* 'name'. The same kind of concord shows up with the word *maga* 'self'. Otherwise Hungarian has no concord or agreement between a preposed attribute and a head noun, nor are coordinated or juxtaposed nouns marked except on the last constituent of the NP. We admit that the inflexion of *ez*, *az*, and *maga* is remarkable and, no doubt, constitutes a counter argument against our original conception of Hungarian. Borin's second argument for allotting Hungarian the status of a true case language, viz. that postposed appositions demand the same ending as the governing noun, we consider more doubtful. In our mind Borin has not convincingly demonstrated that it is the noun proper and not the NP that is determined by the ending. Borin may be right that Hungarian really is an agglutinating case language without agreement, but in as far as the question of suspended affixation is not satisfactorily answered we will remain sceptical.

Summing up, we conclude that noun phrases of natural languages may be organized according to at least three different principles. In one type of language, the most typical representatives of which are Germanic and Slavonic languages, the noun phrase is centred around a single head noun, all determiners usually being specifically marked for syntactic subordination to the head. The difference between Slavonic languages on one hand and a language such as Latin on the other is, accordingly, that the Slavonic ones have two separate sets of case endings for nouns and determiners respectively, whereas Latin has the same set of endings for both word classes. The first type of language would principally allow case marking without agreement. The second type, which includes not only languages such as Latin, Hebrew and Finnish but also modern Romance languages such as Italian and French, has no morphological means to signal the subordination relation between *nomina substantiva* and *nomina adiectiva*. From a syntactic point of

view both types of word within the noun phrase are therefore true nouns. The same holds true of the third type of language, represented by Turkish and Mongolian, but here the noun phrase is an indivisible constituent with regard to sentential syntax. The difference between the three types with respect to structural properties of their noun phrases can be demonstrated graphically as in Figure 1, where A stands for any adjective-like determiner and Infl symbolizes the structural position of the grammatical marker. Thus genuine Type III languages are not proper case languages. They may be flexional languages, but, if so, their prefixes or suffixes function as prepositions or postpositions with respect to the whole NP. Caseless languages can therefore be of two different kinds: on one hand analytical, on the other flexional with grammatical markers determining the whole NP. In addition there may, of course, occur mixed types combining in varying degree the properties of the former and the latter type.

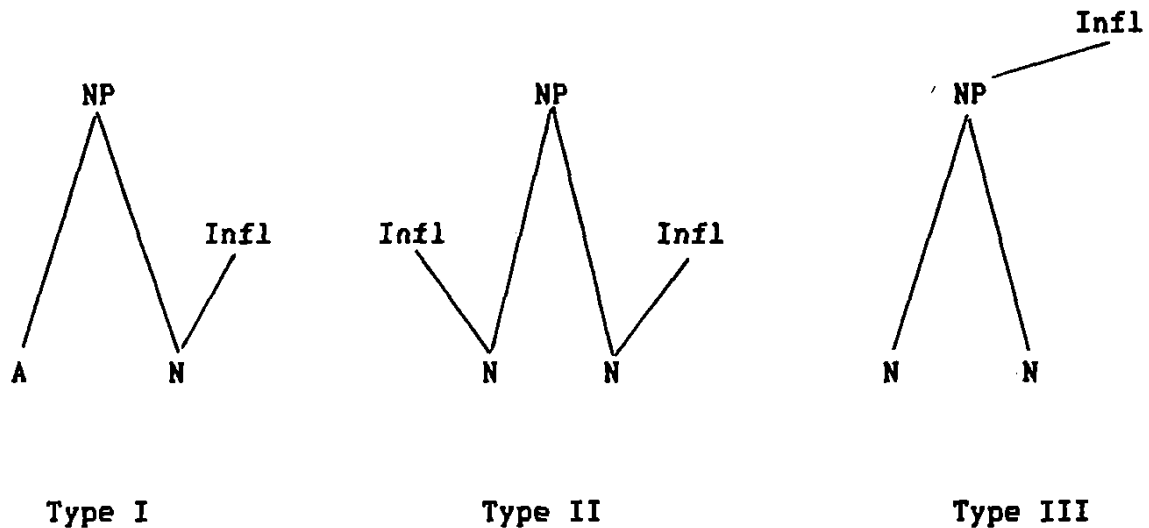


Fig 1

Observe that all language types freely allow other possibilities of signalling the internal structure of the noun phrase. In all types we find subordination relations expressed via embedded prepositional phrases or some kind of casus constructus. Turkish has its *izafet*, but this construction is in fact the same thing as that which is expressed by the possessive particles in modern Dutch. This kind of possessive construction is also typical of

colloquial Norwegian (a filter language) and colloquial German. In colloquial Faroese, which is a full-fledged case language, the genitive has been lost, having been replaced by the same kind of *izafet*. The possessive particles of Finnish also render *izafet*, and yet Finnish is an extremely developed Type II case language with no formal difference between nouns and adjectives. Thus we see that the difference between the three language types is not a matter of the internal structure of the noun phrase as such but rather of the external relation of the noun phrase to other constituents within the sentence. Case markers therefore are attached to noun stems but express relations beyond the scope of the separate nominal constituent.

When discussing the Roumanian definite articles we mentioned in passing that Roumanian has a well developed case system in its pronouns. This feature is also typical of otherwise completely caseless languages, English and Swedish included, though the functioning of the pronominal case forms in these languages seems to be so redundant that they can be disregarded as surface structure phenomena. There are, however, other filter languages, the pronominal case marking of which cannot be so easily neglected. One such language is Bulgarian. It has lost every trace of syntactic case marking in its nouns but nevertheless possesses functioning dative and accusative forms of the personal pronouns just as Roumanian: cf. *Marija dava na Ana jabalka* 'Mary gives Anne an apple' and *tja i ja dava* 'she gives it to her', where *tja* is the nominative, *i* the dative, and *ja* is the accusative singular form of the feminine personal pronoun 'she'.⁴ Such situations are very common among languages all over the world. There has been quite a lot of speculation on the reason or reasons for this phenomenon. Comrie (1981, 179ff.) is inclined to relate it to a universal tendency of special grammatical marking of linguistic entities ranking high with respect to what he refers to under the heading animacy. Just as nouns denoting living beings are more apt to distinguish the nominative and the accusative overtly than neutre nouns, personal pronouns could be considered as ranking higher in animacy than ordinary nouns.

Certainly there is something attractive about Comrie's reasoning. At the same time, however, it is rather dangerous. Reasoning that way one tends to slur or overlook the fact that pronouns and ordinary nouns generally are lexical items of different orders, in spite of the fact that both types of word are used to fill noun phrases. The difference between nouns proper and pronouns is that the former are lexical items mentally connected with prototypical entities that could be referred to as existent

in a conceivable situation even outside a linguistic context, whereas pronouns are exclusively deictic linguistic elements that have no such mental connection but can be used meaningfully only with respect to an already created discourse (cf. Pettersson 1982, 71; Lyons 1977, 660). They are shifters in the sense of Jakobson (1957). There are, in fact, languages which do not differentiate between nouns proper and pronouns in this way. The Japanese word *watasi* is glossed 'I, me', but it is actually an ordinary noun meaning 'that person which is I'. A similar phenomenon is well-known from child language acquisition. At a certain stage children frequently use the second person personal pronoun *you* to refer to themselves. The reason for this is obvious: the child has not yet understood the shifter function of the pronouns; consequently, since he is addressed as *you*, he takes *you* to be a name, an ordinary lexical denomination, or, in other words, a noun.

What has been said so far does not in any way interfere with the claims regarding the relation between pronouns and nouns put forward in Bily (1981, 2, 73-125, 179-215). We still find it unnecessary to differentiate between definite personal pronouns and full NPs as far as coreferential interpretation rules are concerned. From a coreferential point of view all NPs belong to a sliding scale of referential explicitness containing both definite pronouns, semantically empty nouns such as *thingummy*, *gimmick*, *gear* and so on, so-called epithetic anaphoric NPs such as *the bastard* and *the poor man* and full nouns or NPs. Thus the prevailing view of definite personal pronouns as equivalent to bound variables in logic is not linguistically motivated with respect to the interpretation of what they refer to endophorically. The point we want to make in this context, however, is simply that pronouns are radically different from nouns as far as exophoric reference is concerned.

Thus we see that the distinction between nouns and pronouns is not universal in the sense that it is a necessary feature of human language. It would therefore be incorrect to regard pronouns as merely a certain kind of noun. The traditional conception that pronouns are words that fulfill the same functions as nouns without being nouns themselves is no doubt the right one. The distinction between nouns and pronouns adds to the complexity of grammar in such a way that is typical of category split rather than of category subdivision. With all this in mind it is quite natural that we should find syntactic and morphological properties of nouns to be reflected in pronouns, whereas the opposite situation is far from general or even common. On the contrary, langua-

ges tend to develop pronominal systems that in one way or other overtly signal the intrinsic difference between the two word classes. As pointed out above filter languages often have some kind of pronominal case marking while simultaneously being devoid of case marking in nouns. In case languages we often find special morphology as regards pronouns, distinct from that of nouns. Slavonic languages, for example, have accusative marking on pronouns replacing neutre nouns, in spite of the fact that the nominative-accusative distinction is not morphologically signalled in neutre nouns. Finnish has likewise accusative marking of personal pronouns in the singular, while singular nouns have no such marking. Latin developed a specific partitive case in plural pronouns, a case otherwise unknown in the language. Such examples could be multiplied perhaps infinitely. It could very well be that a prerequisite for a language learner to detect the difference between nouns and pronouns is that the language to be acquired possesses some salient features characterizing its pronouns, overtly separating the class of pronouns from that of nouns. Observe that this suggestion in no way challenges earlier explanations such as those referring to the frequency of pronouns in actual speech, making them less vulnerable to analogical change, or those taking into consideration their purport according to parameters such as animacy or definiteness. We are simply pointing out a peculiarity of pronouns that might be a contributing factor, among other things, to the development and retainment of case marking in pronominal systems to the exclusion of case marking on ordinary nouns.

Be this as it may, it is obvious that our decision to define case markers as overt morphemes attached to noun stems is appropriate even with regard to pronouns, since case marking on nouns entails case marking on all kinds of NP heads, whereas case marking in a restricted set of NPs such as personal pronouns does not entail a corresponding case marking on nouns. Consequently, the existence of case marking on pronouns will not as such be a sufficient prerequisite for the classification of a language as a case language. Therefore we are perfectly right in rejecting not only Swedish and English but also Bulgarian and similar languages as representatives of case languages.

One final remark should be added. The fact that we define a case marker as a bound morpheme attached to a noun stem does not mean that we are entitled to recognize all morphemes exhibiting this property to be case markers. Case is a grammatical category by itself, clearly distinguishable from other categories such as number, gender and species. Given the morphological structure of

English alone it is of course impossible to tell whether the ending *s* of the word form *girls* as opposed to *girl* is a marker of number, case, gender, or something else. But with respect to the syntactic and semantic properties of this *s* we can truly state that it is a marker of number. The same thing holds good of the definite ending of nouns in Nordic languages. Since the ending in question actually is added to and thus determines noun stems and not noun phrases, as opposed to Bulgarian and Roumanian, it might be identified as a case marker. Its syntactic and semantic properties, however, are quite different from those of case markers as we have outlined them in Bily & Pettersson (1983; 1984), and thus the Nordic definite article cannot be recognized as a case marker. Accordingly, our criterion is a necessary but not a sufficient condition for distinguishing case morphemes from other morpheme types and case languages from caseless ones. Nevertheless it is a powerful instrument for future research in language typology.

NOTES

1) For reasons of simplicity we will use the term "morpheme" to cover exponents of porte-manteau affixes as well, where case may or may not be an item among others, expressing grammatical properties such as gender, number or species.

2) Data on Roumanian is drawn from Lombard (1973).

3) It should be observed that they are, in fact, true pronouns diachronically.

4) It should, however, be observed that it is impossible to change the order of the oblique pronouns; **tja ja i dava* is thus ungrammatical.

REFERENCES

- Berman, R.A. (1978): *Modern Hebrew structure*. Tel-Aviv: University Publishing Projects
Bily, M. (1981): *Intrasentential pronominalization and Functional Sentence Perspective (in Czech, Russian, and English)* (=Lund Slavonic Monographs 1). Lund

- Bilý, M. & Pettersson, T. (1983): "Towards a theory of morphological case". F. Karlsson (ed.): *Papers from the Seventh Scandinavian Conference of Linguistics*. Helsinki 1982, 568-582
- Bilý, M. & Pettersson, T. (1984): "Kasusteori - en modell för morfologiskt kasus". *Slovo* 28, 5-24
- Bilý, M. & Pettersson, T. (forthcoming): "Ergativity in Dyirbal". *Festschrift for Petr Sgall in the series Linguistic and Literary Studies in Eastern Europe*. Amsterdam: John Benjamin
- Borin, L. (1984): *Är ungerskan ett kasuspråk?*. Stereotyped examination paper, Uppsala University
- Comrie, B. (1981): *Language universals and linguistic typology. Syntax and morphology*. Oxford: Basil Blackwell
- Diderichsen, P. (1946): *Elementær Dansk Grammatik*. København: Gyldendal
- Hjelmslev, L. (1935): *La catégorie des cas. Etude de grammaire générale (=Acta Jutlandica VII)*. Aarhus
- Jakobson, R. (1936): "Beitrag zur allgemeinen Kasuslehre". *Travaux du Cercle Linguistique de Prague*, 6. Prague, 240-288
- Jakobson, R. (1957): *Shifters, verbal categories, and the Russian verb*. Description and Analysis of Contemporary Standard Russian, Department of Slavic Languages and Literatures, Harvard University
- Jakobson, R. (1958): "Morfologičeskie nabljudenija nad slavjenskimi sklonenijem". *American contributions to the Fourth International Congress of Slavists*. Preprint
- Lazard, G. (1957): *Grammaire du persan contemporain*. Paris: Librairie C. Klincksieck
- Levis, G.L. (1967): *Turkish grammar*. Oxford: Clarendon
- Lombard, A. (1973): *Rumänsk grammatik*. Lund: CWK Gleerups bokförlag
- Lyons, J. (1977): *Semantics*. Cambridge: Cambridge University Press
- Mel'čuk, I.A. (1981): "Types de dépendance syntagmatique entre les mot-formes d'une phrase". *Bulletin de la Société de Paris* 76:1, 1981, 1-58
- Peters, L. (1947): *Grammatik der türkischen Sprache*. Berlin: Axel Juncker Verlag
- Pettersson, T. (1982): "Prose and poetry". *Studia Linguistica* 36:1, 64-87
- Pettersson, T. (1983): "Fria adverbial. Divertimento con introduzione". *Nydanske Studier & Almen Kommunikationsteori* 13. København: Akademisk Forlag, 41-74
- Windfuhr, G.L. (1979): *Persian grammar. History and state of its study*. The Hague: Mouton (=Trends in Linguistics. State-of-the-Art Reports 12)