1 Introduction

The notion of person has been widely discussed in many different fields of study including philosophy, sociology, anthropology, psychology, politics, religion, literature and art. Scholars who have addressed the issue of person within these fields have been concerned with questions such as what is a person, who qualifies as a person, what are the cross-cultural differences in the conceptualization of person, what is the relationship between individual identity and person, how do we identify and reidentify someone other than ourselves, when does a person stop being a person, etc. Though the social and cultural construal of personhood is also a topic of concern within linguistics, particularly sociolinguistics and anthropological linguistics, the notion of person in linguistics is primarily conceived of as a grammatical category, on a par with gender, number, case, tense, etc. Accordingly, it is with person as a category of the grammatical system of languages that this book will be primarily concerned.

1.1 Person as a grammatical category

It is often stated that the grammatical category of person covers the expression of the distinction between the speaker of an utterance, the addressee of that utterance and the party talked about that is neither the speaker nor the addressee. The speaker is said to be the first person, the addressee the second person and the party talked about the third person. This, however, is not quite correct. What is missing from the above characterization is the notion of participant or discourse role. In the case of the first and second persons, the grammatical category of person does not simply express the speaker and addressee respectively, but rather the participant or discourse roles of speaker and addressee.[1] The difference between the two characterizations can be appreciated by comparing the personal pronouns *I* and *you* in (1a) with that of the nominals *mummy* and *Johnny* in (1b).

(1) a. I will spank you.
b. Mummy will spank Johnny.

[1] This characterization of the grammatical category person draws on the origin of the term person, i.e. mask. Further, it seeks to provide person forms with a sense as opposed to just a reference.

In certain situational contexts, speakers may refer to themselves and their addressees by their proper names, the relations of kinship that they bear to each other, their titles or occupational roles, etc. Thus in (1b) the word *mummy* could be used by a mother with reference to herself and the name *Johnny* with reference to the child whom she is addressing. In such a case, the words *mummy* and *Johnny* can be said to express the speaker and addressee but they cannot be said to express the discourse roles of speaker and addressee as there is nothing in the words *mummy* and *Johnny* to suggest that they are the speaker and addressee respectively. Conversely, this is precisely what is achieved by the two pronouns *I* and *you* in (1a). *I* is always used to refer to the speaker and you to the addressee.[2] Unlike *mummy* and *Johnny*, the two pronominals cannot have any other referents. Moreover, they do not express anything other than that their referents bear the discourse roles of speaker and addressee respectively. Accordingly, only *I* and *you* and not *mummy* and *Johnny* are expressions of the first and second persons. *Mummy* and *Johnny* are lexical expressions which may be used to refer to the speaker and addressee respectively.

In principle, there is no limit to the nature of the lexical expressions that a speaker may use to refer to herself. By contrast, it would be dysfunctional for languages to have a wide range of expressions to denote the discourse roles of speaker, addressee and third party. And indeed they tend not to. The vast majority of the languages of the world have a closed set of expressions for the identification of the three discourse roles embracing the category of person. The special expressions in question are typically called personal pronouns, or even just pronouns. (The word pronoun without additional qualification is generally interpreted as denoting pronouns expressing person.) In this book, however, we will use the terms *person marker* and *person form* in preference to *pronoun*, as the term pronoun is open to a number of interpretations and even under the most liberal of these, not all grammatical markers of the category person are uncontroversially pronominal. More about the notion of pronoun will be said in section 1.2.

Although the grammatical category of person involves only the three-way distinction of speaker, hearer and third party, this does not mean that languages typically have only three person markers. English, which clearly has many more than three person markers, is by no means exceptional. In fact, despite the array of person markers that English has, it does not qualify as a language rich in person markers. Other languages have many more. For instance, Fijian is said to have as many as 135 person forms. There are also languages with considerably fewer person markers than English. Madurese, an Austronesian language now mainly spoken in Java, has only two, *sengkoq* 'I/me' and *tang* 'my'. For the second and third persons, words meaning 'metaphysical body/spirit' and 'sole/alone' accompanied by a definite marker are used.

[2] This is not quite correct. The second-person form *you* in English, and also in many other languages, has an impersonal or generic use, illustrated in (13b) further below and discussed in more detail in chapter six.

The differences in person-marker inventories found cross-linguistically are in part a reflection of the nature of the grammatical categories in addition to person that the person markers encode. Person markers rarely mark person alone. The grammatical category most closely connected with person is that of number. Two other grammatical distinctions regularly expressed together with person are gender and case. Thus, for example, the English *she* encodes third person, singular number, feminine gender and nominative case, that is the case of the subject. Further grammatical categories which may also be marked together with person include definiteness, obviation, tense, aspect, mood and polarity. The last of these is to be found in the person markers of the Australian language Worora, for example, which, as shown in (2), has a distinctive set of forms used in negative utterances.

(2) Worora (Love 2000:17)

		positive	negative
1SG		ŋaiu	‘ŋaui
2SG		ŋundju	‘ŋungi
3SG[3]	M	‘indja	‘kaui
	F	‘nijina	‘njuŋgi
	NT	‘wuna	‘kui
	NT	‘mana	‘maui

In addition to other grammatical categories, person markers may also encode information pertaining to their referents, for example, the social status of the referent vis-à-vis the speaker, their location relative to the speaker or addressee or, much more rarely, their kin relationship and/or generation level. A celebrated instance of person forms reflecting generation levels comes from Lardil (Hale 1966), another Australian language. In Lardil, in the dual and plural, one set of person forms is used for persons who belong to the same generation level or are two levels apart, and a different set of forms for persons one or three generations apart. Thus the form of the second-person dual ‘you two’ when used to refer to, say, a brother and sister or a grandparent and their grandchild is *kirri*, but when used to refer to a parent and child or great-grandparent and their grandchild is *nyiinki*. More complex systems involving not only considerations of generation level but also of membership within a given moiety (i.e. a particular set of kin categories) are found in other Australian languages, such as Arabana-Wangkangurru (Hercus 1994:117), Adnyamathanha and Kuyani (Schebeck 1973). In these last two languages there are twelve different sets of person markers to mark the kinship associations of the people to whom the person forms refer and, in some instances, also the speaker’s relationship to these people. In contrast to the Australian languages mentioned above, in the Tibeto-Burman language Dhimal (King 2001) there are special person forms just for the first- and second-person singular which are reciprocally used only between two distinct groups, one being the parents of

[3] Membership in the two sets of neuter forms in Worora, the *wuna* set and the *mana* set is lexically determined.

a husband and a wife and the other, a man and his wife's senior relatives. In exchanges between these two groups the first-person singular is *kya* which contrasts with the typical *ka*, and the second-person singular is *nya* rather than *na*. Yet another factor, in part relating to referents, which has been noted to be encoded in the person markers of a language is speech style. Jacquesson (2001:123) reports that in several dialects of Tiddim, a Tibeto-Burman language, there are two sets of verbal person markers for all three persons: a prefixal set and a suffixal set. The former is used in narratives, the latter in everyday speech.

The other major source of differences in person-marker inventories is variation in morpho-phonological form. In some languages all the person markers are independent words, while others, in addition to such forms, also have person clitics and/or affixes and/or covert, that is zero forms. Bulgarian, for example, apart from independent forms, which may be used for all syntactic functions, has clitics used for objects, and affixes (fused with tense/aspect) used for subjects. All three forms occur in (3).

(3) Bulgarian (Dimitrova-Vulchanova & Hellan 1999:490)
Na Ivana kniga-ta az mu= ja= dadox
to Ivan book-DEF I 3SG:DAT 3SG:ACC give:1SG
'I gave the book to Ivan.'

1.1.1 Person paradigms

The person markers found in languages do not occur in isolation but rather in closed sets called paradigms. Simplifying somewhat, a paradigm is a set of linguistic expressions that occur in the same syntactic slot in the language. Moreover, each member of a paradigm is in complementary distribution with every other member of the same paradigm. Thus the English person forms *I/you/he/she/it/we/you/they* constitute one paradigm, as each may occur as the subject of an utterance and the use of any one form excludes the possibility of using any of the others (apart from coordinations). The person forms *me/you/him/her/it/us/you/them* belong to another paradigm, since they are employed as objects and complements of prepositions but, crucially, not as subjects. And the forms *my/your/his/her/its/our/your/their* make up a third paradigm used as attributive possessors. In addition to performing the same syntactic function, the members of a single paradigm are also assumed to have the same morpho-phonological form, that is to be all independent forms, or all clitics or affixes, etc.[4] Consequently, independent and clitic forms, such as the Bulgarian third-person masculine object forms *jemu* and *mu*, are seen as belonging to two different paradigms.

While there are languages which have only one paradigm of person markers used for all syntactic and discourse functions, most languages have several. An

[4] Occasionally differences in morpho-phonological form are found within what is considered to be a paradigm. For instance, in Fur (Jakobi 1990:28), the dependent object person markers in the singular are suffixes, in the plural clitics or weak forms.

important point to remember is that in those languages which have more than one paradigm of person markers, the structure of the different paradigms need not be the same. The paradigms may differ in regard to the person, number, gender and other distinctions marked. A particularly clear example of such differences between paradigms is presented in (4) from Vinmavis, an Oceanic language spoken on the island of Malakula, in which the independent person markers evince a singular/plural distinction and an inclusive/exclusive one (see section 3.2.1), while the subject prefixes exhibit an opposition between the singular, dual and plural but no inclusive/exclusive contrast.

(4) Vinmavis (Crowley 2002b:640, 644)

Indep form		Subject prefix (non-future)	
1SG	no	1SG	nV- /na-
2SG	gu	2SG	u-
3SG	i	3SG	i-
1PL INCL	get	1DU	er-
1PL EXCL	gemem	2DU	ar-
2PL	gem	3DU	ar-
3PL	ar	1PL	it-
		2PL	at-
		3PL	at-

Although the number of distinct person forms in the two paradigms is actually the same, seven (due to the homophony between the second and third persons in both the dual and plural, in the case of the subject prefixes), they differ radically in their internal structure. The existence of such differences makes it difficult to discuss the person system of a language as a whole. Linguists are often tempted to make general statements about the nature of person marking in a language. Such general statements, however, are possible only for some languages, but definitely not others. This has to be kept in mind while reading this book. Just because a particular language is cited as displaying a particular property or feature in some person paradigm, this does not mean that the same holds for all the person paradigms.

1.1.2 First and second persons vs third person

It is generally acknowledged that "there is a fundamental, and ineradicable, difference between the first and second person, on the one hand, and the third person on the other" (Lyons 1977:638). One manifestation of this difference is that whereas the first and second persons are regularly referred to essentially only by person markers, reference to the third person can be achieved by any lexical expression. It should therefore be unsurprising that languages may have first- and second-person markers but no third-person ones. In many of the languages which lack person markers for the third person, demonstrative pronouns corresponding to the English *this* and *that* are used in lieu of third-person markers. This is the case, for instance, in Basque, Comanche, Imbabura Quechua, Lak,

Lavukaleve and Maricopa. In other languages reference to the third person is achieved only via full nominal expressions. This appears to be so, for instance, in Salt Yui (Irwin 1974:32), a Papuan language, where third person is indicated by a specific noun, such as *yai* 'male' or *al* 'female' followed by a demonstrative. There is no distinction in number. Thus *yai i* denotes 'he' or 'they masculine' and *al i*, 'she' or 'they feminine'. And in yet other languages either full nominal expressions are used or, alternatively, no overt expression at all, the absence of an overt expression being interpreted as denoting third person.

A difference between first and second persons as opposed to the third may also be manifested in languages which have person markers for all three persons. Often the forms of the first and second persons are quite different from that of the third. As (5) illustrates, this may be observed in Nosu, a Tibeto-Burman language belonging to the Northern Yi group, spoken by over two million people in Sichuan and northern Yunnan, China.

(5) Nosu – Northern Yi (Bradley 1993:185)
1SG ŋa^{33}
2SG nɯ a^{33}
3SG ts^{h}z^{33}

There may also be a difference in the order of third-person forms as compared to that of first- and second-person forms. For example, in Takale and Gamale, two dialects of the Tibeto-Burman language Kham (Watters 1993:105), when the agent is first or second person, the agent forms precede the patient forms. But when the agent is third person, the agent forms follow the patient ones. Compare (6a,b) with (6c).

(6) Gamale (Watters 1993:107)

a. Nə-hnə-kəŋ-khě
2SG(A)-look-1SG(P)-PAST
'You looked at me.'

b. Ye-hnə-rə
1SG(A)-look-3PL(P)
'I looked at them.'

c. Ya-hnə-kəŋ-wo
PAST-look-1SG(P)-3SG(A)
'He looked at me.'

Another, not uncommon, difference between first and second persons as opposed to third person involves case marking. Third-person forms may take a different set of case markers than first- and second-person forms. For instance, in the Australian language Wambaya (Nordlinger 1998), there are three separate third-person forms, one for the S (sole argument of an intransitive clause), another for the A (agentive argument of a transitive clause) and a third for the P (patient-like argument of a transitive clause). But there are only two forms for the first and

second persons, one for the S and A, and another for the P. Particularly frequent are differences between the first and second persons as compared to the third in regard to number and gender. Number distinctions are often neutralized in the third person, while gender is rarely manifested by second- and hardly ever by first-person forms. (This is discussed in detail in chapter three.)

All of the above differences are typically seen to be a consequence of the fact that first- and second-person forms are inherently deictic expressions, that is their interpretation is dependent on the properties of the extralinguistic context of the utterance in which they occur. Although the first person is always the speaker of the utterance and the second the hearer, the actual identity of each depends on who utters the utterance that contains them to whom, when and where. They belong to the class of expressions often referred to as shifters (Jakobson 1971). Third-person forms, on the other hand, are essentially anaphoric expressions. Their interpretation depends not on the extralinguistic but on the linguistic context of the utterance.[5] The referent of *he* or *she* is typically established by the preceding discourse, as in (7) or, less often, by the following discourse as in (8).

(7) – There's no sign of John.
– He must have missed his train again.

(8) – She is late again.
– You mean Sally.
– Yes.

Third-person forms may be used deictically, as when someone says (pointing to a grinning child who has just been given an enormous ice cream) *He's happy*. Their anaphoric use is, however, the basic one. In fact in some languages, third-person forms can only be used anaphorically, deictic reference being achieved via demonstratives. Much less frequently, in addition to demonstratives there are two sets of third-person forms, one set for deictic reference, and another for anaphoric. This is so in Udihe (Nikolaeva & Tolskaya 2001:753–4), a Tungusic language spoken by about a hundred people in the Russian Far East.

In the linguistic literature, mention is sometimes made not only of the first, second and third persons but of a fourth person. This label is applied to several quite different kinds of categories. For instance, in the French grammatical tradition the term fourth person is often used for the first person plural. In Amerindian studies, especially of Algonkian languages, the label fourth person is used with reference to a less important third person, called an obviative as opposed to a proximate. And in discussions of anaphoric relations across clauses, the term fourth person is used for special third-person forms that indicate coreference, which are also termed logophoric or long-distance reflexives. Under none of the above uses does the fourth person qualify as a bonafide additional discourse category. Therefore, I see no reason for using the term here.

[5] In place of the terms deictic and anaphoric, some linguists use the terms exophoric and endophoric.

Whereas some scholars seek to expand the number of categories comprising person from three to four, others seek to reduce it. As mentioned earlier, of the three persons only the first and second persons are actual participants in the speech act realized by the utterance containing them. The third person is a not a participant of the speech act. Some linguists, most notably Benveniste (1971), argue that the grammatical category of person should therefore be seen as embracing only the first and second persons with the third person being a non-person. This is not the view adopted in this book. While fully acknowledging the distinctive nature of the third person relative to the first and second, I see no advantage in excluding the third person from the category of person, particularly in a cross-linguistic study such as this one. In fact, as will become apparent in the course of our discussion, doing so would severely skew our understanding of a number of facets of the category of person.

1.2 The universality of person markers

Despite statements such as the following by Benveniste (1971:225) "A language without the expression of person cannot be imagined", the universality of person as a grammatical category is sometimes called into question. The issue of whether all languages display the grammatical category of person is inherently tied to the issue of whether all languages have the category of personal pronoun. What constitutes a personal pronoun is in turn a matter of considerable controversy. The notion of pronominality has been and continues to be discussed in several different contexts and thus the features taken to be characteristic of pronouns are very much dependent on what they are being compared with or opposed to. Traditionally, personal pronouns have been opposed to nouns or NPs. Within the generative approach, ever since Chomsky's (1981) binding conditions, they have been contrasted primarily with anaphors (reflexives). Another line of inquiry opposes personal pronouns to person agreement markers (e.g. Bresnan & Mchombo 1987). And yet other studies seek to characterize pronouns in terms of a scale of structural deficiency (Cardinaletti & Starke 1999). We will have cause to consider all of the above at various points in our discussion, but for the time being, let us just concentrate on the pronoun vs noun distinction.

Traditionally, a personal pronoun is taken to be a morpho-syntactic category, which may be used to substitute for nouns or rather NPs, but differing from the latter in its morphological and syntactic properties. Under this traditional approach various languages, most notably South-east Asian languages such as Thai, Burmese, Vietnamese and Japanese, have been argued to lack personal pronouns, since the expressions used to indicate person display properties of nouns.[6]

[6] Other languages, such as the Salishan Northern Straits Salish (Jelinek 1998) and Halkomelem (Wiltschko 2002) have been argued to possess only bound pronouns. Such languages will be discussed in section 2.1.1.

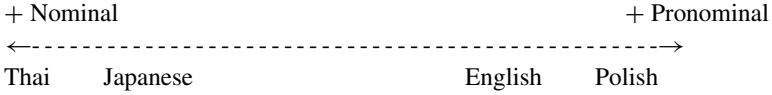
More recently, however, what constitutes a pronoun has come to be viewed somewhat differently. In the generative literature (e.g. Noguchi 1997, Bresnan 2001b) a pronoun is seen to be not a morpho-syntactic category but rather a feature that sets off certain lexical items from others. The relevant feature is referential dependency; although pronouns are used to refer to individuals and entities, the identity of their referents can be determined only by the extralinguistic context (for first- and second-person forms) or typically the linguistic context (for third-person forms) or inferentially. This referential deficiency distinguishes them from both proper nouns, which are capable of identifying a referent by themselves, and common nouns, which are semantic predicates requiring a determiner to enable them to be used as referential expressions. In terms of this approach, all or some of the South-East Asian languages mentioned above are seen to have pronouns, but differing in syntactic category from the pronouns in, say, English. English pronouns are treated as determiners, and Japanese pronouns as nouns. The morpho-syntactic differences between the relevant forms in the two languages are thus seen to follow from differences in their categorial status but not in their pronominal status.

In the functional literature, in turn, pronouns in the main continue to be viewed as a morpho-syntactic category but often the distinction between pronoun and noun is considered to be not discrete but scalar, with some pronouns exhibiting less prototypically pronominal and more nominal characteristics than others. This position is most clearly articulated by Sugamoto (1989), who posits the characteristics in (9) as representing the pronominal extreme of what she calls the pronominality scale:

(9) a. closed class membership
b. lack of morphological constancy
c. lack of specific semantic content
d. lack of stylistic and sociolinguistic implicative properties
e. expression of grammatical person
f. inability to take modifiers
g. restrictions on reference interpretation

These criteria can be used to place person markers on a pronominality scale both across languages and also within languages. For example, if applied to the personal pronouns in English, Polish, Japanese and Thai, the Polish personal pronouns emerge as more pronominal than the English, both as considerably more pronominal than the Japanese forms, and the Japanese forms as more pronominal than those in Thai, as exemplified on the pronominality scale in (10).

(10) The pronominality scale



Let us first consider the English personal pronouns.

English personal pronouns have most of the properties in (10). They belong to a closed class and, unlike nouns, are not morphologically transparent as far as number or case is concerned. Whereas number with most nouns is indicated by suffixation of /-(ə)z/ (e.g. *dog* vs *dogs* or *dress* vs *dresses*), with pronouns it is indicated by suppletion of the stem (e.g. *I* vs *we*). And whereas nouns may be marked for the genitive case by /'s/ (e.g. *mother's friend*), pronouns again have separate forms (e.g. *I* vs *my*). Further, pronouns convey no semantic content other than that of the grammatical features which are associated with them and do not vary stylistically, while nouns may do so (e.g. *mother* vs *mummy* vs *mum*). And clearly pronouns distinguish between the first, second and third persons, while nouns are necessarily third person. English personal pronouns can, however, co-occur with some of the modifiers that are found with nouns. The plural forms may be modified by a low numeral (e.g. *us two*, *we four*), the accusative forms may be modified by certain adjectives, such as *poor*, *kind*, *evil*, *lucky*, (e.g. *poor me*, *lucky you*) and the nominative forms may be modified by a non-restrictive relative clause, as in *I, who have nothing, he who strives, wins*. As for reference, the personal pronouns are clearly restricted in regard to their referential interpretations in the sense outlined above. While nouns may be used for both definite (e.g. *the book*, *this book*, *my book*) and indefinite reference (e.g. *a book*, *some book*), personal pronouns are (with few exceptions) definite.[7] This is reflected in the fact that they do not normally occur with any of the determiners, i.e. articles (e.g. **the he*), demonstratives (e.g. **this she*) or genitives (**my he*) which transform a noun into a definite referential expression and are normally incompatible with the indefinite article (e.g. **a she*). The qualification normally is necessary in view of examples such as those in (11), taken from Noguchi (1997:778–9).

(11) a. This is not the real me.
b. Do you know the real you?
c. That's not a he; that's a she.
d. It's a he!

The use of the definite article with personal pronouns as in (11a,b) is highly restricted; for most speakers the personal pronoun must be the accusative singular form and, for some, even just the first and second persons (?*This is not the real him*.), and the adjective must be present (**This is not the me*.). The indefinite article is possible only with the third-person nominative (**It's a him*). Such usage thus cannot be seen as actually undermining the essentially definite nature of the personal pronouns.[8] The above notwithstanding, English personal pronouns are not always used strictly referentially, that is to refer to concrete entities or individuals. For example, in (12) *Kate* has no specific person in mind and thus *he* is used non-referentially.

[7] For a discussion of reference and definiteness see, e.g., Lyons (1977:177) or Allan (2001:59, 440). The issue will be resumed in sections 4.1 and 4.3.4.
[8] One way in which such atypical co-occurrences of the article and personal pronoun are dealt with is by assuming that a category conversion has taken place, from a pronoun to a noun.