

CHAPTER 2

Bi-directional vs. uni-directional asymmetries in the encoding of semantic distinctions in free and bound person forms

Anna Siewierska
Lancaster University

Dik Bakker
University of Amsterdam

Abstract

In languages in which the semantic distinctions encoded on free and bound person forms are not identical, the more elaborate semantic distinctions may be borne either by the free forms or the bound. The present chapter explores two hypotheses which have been advanced to account for the existence of bi-directional asymmetries in the degree of elaboration of the semantic distinctions encoded in free as opposed to bound person forms, both framed in the context of the typological distinction between free-pronoun and bound-pronoun languages suggested by Bhat (2004). The first hypothesis associates the more elaborate semantic distinctions of whatever type with free pronouns in free-pronoun languages and bound pronouns in bound-pronoun languages. The second hypothesis distinguishes between the semantic dimensions of person and clusivity on the one hand and number and gender on the other. The former are seen to favour greater elaboration on whatever person forms are the primary exponents of the category person in a language, the latter to favour the secondary person forms. The investigation, carried out on the basis of a comparison of the semantic distinctions encoded in free and bound person forms in a cross-linguistic sample of 328 languages, lends relatively little support to either of the two hypotheses. The results are shown to be in the main compatible with the tenets of grammaticalization.

1. Introduction

The vast majority of the world's languages have both free and bound person forms. More often than not the semantic distinctions encoded in the two sets of forms are exactly the same. This holds with respect to the grammatical categories expressed, i.e. person and typically also number and often also gender and inclusivity. It also holds for the degree of elaboration of these categories, i.e. the variety of oppositions encoded, for example, in the case of number, singular vs. plural or singular vs. dual

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

vs. plural or, in the case of gender, masculine vs. feminine only in the third person singular or masculine vs. no masculine in the third person singular and plural etc. There are, nonetheless, quite a few languages in which the semantic distinctions encoded in free and bound person forms are not identical. In some, the semantic distinctions encoded in the free forms are more elaborate than those in the bound. This is the case, for instance, in the Australian language Ungarinjin with respect to the grammatical category number. As shown in (1) below the free person forms evince a four way number distinction of singular vs. dual vs. paucal vs. plural, while the bound prefixes of the intransitive S have only a singular vs. plural contrast.

(1) Ungarinjin (Rumsey 1982)

	Free	S
1SG	ɲjin	ɲa-
2SG	njaŋan	njin-
1DU INCL	ɲarun-njiri	—
1DU EXCL	njarun-njiri	—
2DU	nurun-njiri	—
1PAUC INCL	ɲarun-njina	—
1PAUC EXCL	njarun-njina	—
2pauc	nurun-njina	—
1PL INCL	ɲarun	ɲar-
1PL EXCL	njarun	njar-
2PL	nurun	gur-

In other languages, conversely, the semantic distinctions encoded in the bound forms are more elaborate than those in the free. Such a situation with respect to again number obtains in the Austronesian language Sye. We see in (2) that the S prefixes (the forms given are those in the recent past) manifest a dual in the first person which is lacking in the free forms.

(2) Sye (Crowley 1998: 41, 90)

	Free	S
1SG	yau	yaco-
2SG	kik	ko-
3SG	iyi	co-
1DU INCL	—	koku-
1DU EXCL	—	kaku-
1PL INCL	koh	kokli-
1PL EXCL	kam	kakli-
2PL	kimi	ku-
3PL	iror	cu-

There are even languages in which the free forms exhibit a higher degree of elaboration than the bound with respect to one semantic dimension, while the bound exhibit a higher degree of elaboration than the free in regard to another semantic dimension. We can observe this, for example, in certain dialects of the Papuan language Senatani in which the free forms have an inclusive–exclusive contrast which is lacking in the bound forms, while the bound forms manifest a dual number lacking in the free forms.

(3) Sentani (Cowan 1965: 16, 28)

	Free	S
1SG	da	a
2SG	wa	-(j)é
3SG	na	-Ø-w
		1DU -en, e(j)
		2DU -ew
		3DU -ej
1INCL	(e)	
1EXCL	me	1PL -an -a(j)
2PL	ma	2PL -aw
3PL	na	3PL -aj

In the light of the above, two questions pertaining to the distribution of semantic distinctions in free and bound person forms suggest themselves. First of all, is it at all possible to make any generalizations with respect to the directionality of the asymmetries in the degree of elaboration of the semantic distinctions encoded in free as opposed to bound person forms? For example, are some types of languages more likely to manifest more elaborate semantic distinctions in free person forms while others in bound person forms? And secondly, to what extent do the asymmetries in degree of semantic elaboration of free and bound forms depend on the semantic distinction in question?

The present chapter seeks to provide answers to these questions by comparing the semantic distinctions encoded in free and bound person forms in a cross-linguistic sample of 328 languages containing both free person forms and bound person forms (for the S). This sample is drawn from a larger computerized data base compiled by the authors over the last ten years available at <<http://www-uilots.let.uu.nl/ltrc/agreement.htm>>. The investigation takes as its point of departure two hypotheses which have been entertained in the literature, though not always explicitly articulated as such. Both hypotheses assume that the existing asymmetries in the distribution of semantic distinctions in free and bound person forms are bi-directional rather than uni-directional, i.e. in some instances they favour free forms, in other instances bound forms. Underlying both hypotheses is the not uncontrovers-

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

sial assumption that languages differ with respect to which person forms, the free or the bound, may be seen as the primary exponents of the grammatical category person (see especially Bhat 2004: 15–18).¹ Let us for the time being assume a bifurcation of languages into those in which the free forms are the primary exponents of person, which may be referred to as free pronoun languages, and those in which the bound forms are the primary exponents of person, which may be referred to as bound pronoun languages. The first hypothesis to be examined is: if the semantic distinctions encoded in free and bound person forms of a language are not identical, the more elaborate semantic encoding will be expressed by the primary exponents of the grammatical category person in that language. This entails that in free pronoun languages the more elaborate semantic distinctions will be borne by the free forms as opposed to the bound forms, and in bound pronoun languages the more elaborate semantic encoding will be borne by the bound forms as opposed to the free forms.

The second hypothesis is a more refined version of the first in that it takes additional account of the function of the semantic distinctions in question.² Of the major four semantic distinctions encoded in person forms mentioned above, person and inclusivity relate directly to the identification of the speech roles. Person encodes the speech roles of speaker (the first person), hearer (second person) and other (third person). And inclusivity encodes various combinations of the speech roles, speaker and hearer (1+2) or speaker, hearer and other (1+2+3) or speaker and other (1+3) etc. Number and gender, on the other hand, have more to do with the properties of the discourse referents fulfilling the speech roles in question, either with their cardinality or sex, humanness or animacy. Since the major function of person forms is typically assumed to be the encoding of the distinction between the three persons and especially between the first and second person, rather than the encoding of the properties of the referents of these speech roles, the primary exponents of the category of person in a language may be expected to be particularly well differentiated with respect to the dimensions of person and inclusivity but less well differentiated with respect to number and gender. Accordingly, the second hypothesis is: if the semantic distinctions encoded in free and bound person forms of a language are not identical, the more elaborate person and inclusivity distinctions will be expressed by the primary exponents of the grammatical category person in that language, the more elaborate number and gender distinctions by the secondary exponents of person. In other words, in free pronoun languages greater elaboration of person and/or inclusivity distinctions will be associated with the free forms, greater elaboration of number and/or gender with the bound forms. And conversely, in bound pronoun languages; greater elaboration of person and/or inclusivity will be associated with bound pronouns, while greater elaboration of gender and/or number with free forms.

Both of the hypotheses stated above, and particularly hypothesis two, are somewhat at odds with the widely accepted claim that bound person forms originate from free person forms in the process of grammaticalization (see e.g. Givón 1976; Lehmann 1982b; Corbett 1991; Siewierska 1999). The grammaticalization scenario entails that accompanying the loss of independent status and phonological reduction of forms is semantic reduction or bleaching (see e.g. Lehmann 1982a: 236; Haspelmath 1998: 1050; Croft 2000).³ Accordingly, one would expect bound person forms to be either equally or less differentiated semantically than the free forms irrespective of language type. Yet hypothesis one predicts equal or greater semantic differentiation of bound forms than of free forms in bound pronoun languages, and hypothesis two in both bound pronoun and free pronoun languages, though with respect to different semantic dimensions. While neither of the two hypotheses is actually irreconcilable with the grammaticalization scenario, the corroboration of either would raise the question of the mechanisms leading to the bi-directional as opposed to unidirectional asymmetries in the semantic distinctions encoded in person forms. An investigation of the two hypothesis is thus of interest not only for typological reasons but also for diachronic ones.

The chapter will be organised as follows. Section two presents the person forms covered by this investigation, the semantic dimensions which they encode and the way we have determined the degree of elaboration of each of these dimensions in a paradigm. Section three elaborates on the classification of languages in terms of which person forms are the primary exponents of the category person and presents the typology that we have used in our analysis. Section four is devoted to testing the bi-directionality of the asymmetries in semantic distinctions encoded in free and bound person forms as captured in the two hypotheses presented above on the basis of the languages in our sample. In section five we briefly consider how our findings can be accommodated within the context of the grammaticalization scenario. Finally, section six provides some concluding remarks.

2. Semantic distinctions in free and bound forms

The distinction between free vs. bound person forms, though so frequently invoked is not uncontroversial. So before we proceed to consider the semantic distinctions in person forms, a few words about our classification of free and bound forms are in order.

Typically, what is meant by an free person form (or its equivalents such as independent, full, disjunctive etc.) is a person marker which constitutes a separate word and may take primary word stress, such as the English *I, me, you, she, they*. Word status in turn is associated with properties such as the ability to be involved in coor-

dinations, the possibility of being deleted under appropriate discourse conditions and the possibility of being modified by another word (see e.g. Zwicky 1985; Dixon and Aikhenvald 2002). Most languages have at least one paradigm of person forms which qualify as free in the above sense and many languages have several such paradigms. For the purpose of this investigation we have selected from each language the free nominative or absolutive forms or, in the case of languages lacking such forms, the forms that may be used as single word answers to questions such as “Who is coming?”. The selected forms are thus not homogenous from either the syntactic or pragmatic point of view. They do, however, all satisfy some subset of the commonly used criteria for word status and may carry the pragmatic function of focus.

Turning to the notion bound, we have subsumed under this notion what others call dependent or reduced person forms which include so-called weak forms, clitics and affixes.⁴ In terms of function, the forms in question correspond to what in the wake of Bresnan and Mchombo (1987) are referred to as anaphoric and grammatical agreement markers. As for their syntactic relation, we will restrict our attention to bound person forms used in intransitive clauses featuring unergative predicates such as those illustrated in (4), (5) and (6) below.⁵

- (4) Woleaian (Sohn 1975: 151)
 (Gaami) gai lag!
 you:PL 2PL go
 ‘You(pl) go!’
- (5) Marubo (Romankevicius Costa 1998: 66)
 (Ia-Ø) in=wi’ša-i-ki.
 I:ABS 1SG-write-PRES
 ‘I am writing.’
- (6) Cora (Casad 1984: 178)
 N^vε-yaana.
 1SG-smoke
 ‘I’m smoking.’

Although person forms may encode semantic dimensions other than person, number, inclusivity or gender, our comparison of the free and bound forms is restricted to these four dimensions. Each of these four semantic dimensions may be elaborated to a greater or lesser degree within a person paradigm, including person itself. Therefore in comparing free person forms and their corresponding bound counterparts we have taken into account not only whether a given semantic dimension is present or not but also the extent to which it is manifested in a paradigm. In particular, we have taken into consideration the number of oppositions expressed within each dimension, their distribution within the paradigm and the major pat-

terms of homophonies displayed. How exactly we have assigned weights along each semantic dimension is explicated below.

2.1 Person

Of the three persons typically taken to comprise the grammatical category person, the first and second person are generally viewed as being more central to the category than the third. Moreover, the first person is typically ranked over the second. In measuring the degree of semantic elaboration of person paradigms we have assumed the above ranking, i.e. we have used the traditional person hierarchy in (7).

$$(7) 1 > 2 > 3$$

Within the three person system, a paradigm fully specified for person is one in which each of the three persons is distinguished from the other by a phonologically distinct form.⁶ Departures from full specification may involve absence of one of the three persons altogether or the presence of homophonies.

Paradigms lacking a particular person distinction are not very common, particularly in the case of free forms.⁷ Nonetheless, there are languages which have free forms only for a subset of the three persons, typically for the first and second person, the third being rendered by lexical expressions. This is the case, for example, 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. Bound forms may also lack one (or more) of the three persons. For example, the Mixtecan language Copala Trique (Hollenbach 1992: 349) has subject clitics for the first and third person singular but not for the second person singular or any person in the non-singular apart from a generalized inclusive. The person-number combinations lacking clitics are expressed by free forms. To give another example, in the Brazilian language Trumai (Guiradello 1999) there are absolutive clitics for the third person but none for either the first or the second, which again are expressed by free forms. And in Noni (Hyman 1981: 77) a Grassfields language of Cameroon, a bound person marker in the form of a homorganic nasal obligatorily occurs on the verb but only for the first person singular, no other person-number combination. Person paradigms which lack a particular person or person-number category such as the above may be referred to as defective. In assessing their degree of defectiveness with respect to person, we have applied the ranking of the three persons in the person hierarchy in (7). Accordingly, absence of the first person has been weighted more severely than of the second, and of the second more severely than of the third.

Underspecification for person due to homophony is far more common than are defective paradigms. Homophonies may involve any two persons, as in (8a, b, c) or even all three, as in (8d).

- (8) a. 1 vs 2 = 3
 b. 1 = 3 vs. 2
 c. 1 = 2 vs. 3
 d. 1 = 2 = 3

Pattern (8a), homophony between the second and third person with a distinct form for the first can be observed in the Papuan language Wambon in its subject suffixes as illustrated in (9).

- (9) Wambon (De Vries 1989: 22)
- | | | | |
|-----|-----|-----|----------------------------------|
| 1SG | -ep | 1PL | -eva |
| 2SG | Ø | 2PL | -e (past)/-na(future)/Ø(present) |
| 3SG | Ø | 3PL | -e (past)/-na(future)/Ø(present) |

The second pattern (8b), where the first and third persons are homophonous and distinguished from the second, is illustrated in (10) below on the basis of the singular S/A suffixes of the realis mood in another Papuan language, Koiari.

- (10) Koiari (Dutton 1996: 23)
- | | Present | Past |
|-----|---------|------|
| 1SG | -ma | -nu |
| 2SG | -a | -nua |
| 3SG | -ma | -nu |
| 1PL | -a | -nua |
| 2PL | -a | -nua |
| 3PL | -a | -nua |

Pattern (8c), homophony between the first and second person and a distinct form for the third, is shown in (11) from yet another Papuan language, Au.

- (11) Au (Foley 1986: 73–4)
- | | | | | | |
|---------|----|-------|----|---------|----|
| 1SG | h- | 1DU | w- | 1PL | m- |
| 2SG | h- | 2DU | y- | 2PL | y- |
| 3SG M/N | k- | 3DU M | t- | 3PL M/F | n- |
| 3SG F | w- | 3DU F | n- | 3PL N | m- |
| | | 3DU N | m- | | |

Pattern (8d), homophony of all three persons, is conditional on there being actual person distinctions in some other person–number combinations of the same paradigm. Such is the case in Koiari, which, as (10) demonstrates, has a 13 vs. 2 opposition in the singular and a three way homophony in the plural.

While homophonies involving non-distinctiveness of all three persons obviously constitute a more radical departure from a fully specified person paradigm than

those involving non-distinctiveness of two persons, the latter can be further differentiated with reference to the person hierarchy in (7). Although even finer grained distinctions could be made, we have differentiated between homophonies which obliterate the distinction between the first and second person as in (8c) as opposed to ones which do not as in (8a) and (8b).

We have sought to capture the various types of underspecification for person in a four-point scale. The maximum 4 points has been assigned to a paradigm fully specified for person, 3 points to one defective with respect to the third person, 2 points to paradigms involving a homophony between other than the first and second persons and 1 point to paradigms defective in the second person or the second and third person. Paradigms manifesting homophony between first and second person or defective for both first and second person as well as those manifesting a three-way homophony have been assigned zero points.

2.2 Number

Although there are person paradigms which do not manifest the semantic dimension of number, the overwhelming majority do. Most commonly person paradigms display only a two way number opposition, of singular, denoting exactly one first, second or third person and non-singular, denoting more than one individual. There are, however, person paradigms which also feature a dual as in Sye, Sentani and Au, illustrated earlier above in (2), (3) and (10) respectively. Yet higher numbers, namely trial as in the Austronesian language Larike (12) or paucal (several or a few) as in Southeastern Ambrym (13) are also to be found in person paradigms, though so far they have been attested mainly among the languages of Micronesia.⁸

(12) Larike (Laidig 1993: 321)

SG	DU	TRIAL	PL	
1INCL	itua-	itidi-	ite-	
1EXCL	au	arua-	aridu-	ami-
2	a-/ai-	irua-	iridu-	imi-
3HUM	mati-			
3NHUM	i-			iri-

(13) Southeastern Ambrym (Crowley 2004: 666)

SG	DU	PAUC	PL	
1INCL		rali-	rati-	ri-
1EXCL	ni-	mali-	mati-	mu-
2	ui-	muli-	muti-	mu-
3	i-	lali-	lati-	li-

The cross-linguistic distribution of number oppositions in person paradigms is seen as conforming to the number hierarchy in (14), i.e. in the main the presence of a trial or paucal implies the presence of a dual, that of a dual a plural, and that of a plural, a singular (see Corbett 2000: 38).

- (14) singular > plural > dual > trial/paucal

Assuming that the unmarked situation is the presence of the same number oppositions, be it singular vs. non-singular or singular vs. dual vs. plural etc. with all three persons, departures from this norm may be treated as instances of homophony, though horizontal as opposed to vertical homophony.

Our assessment of the degree of elaboration of number in a paradigm is based on a six point scale. The maximum of 6 points has been assigned to paradigms manifesting a four-way number opposition, 4 points to paradigms with a three-way number opposition, 2 points to paradigms with a two-way number opposition, 1 point to paradigms with only singular number and 0 points to paradigms lacking the dimension of number. Paradigms featuring a horizontal homophony have been penalized by having a point subtracted. Thus, for example, a paradigm with a four-way number opposition and horizontal homophonies has received 5 points rather than 6. We have not taken into account in the point system the type of horizontal homophony present, be it with respect to the person involved or the number oppositions embraced. The former will, however, be taken into consideration in the discussion in Section 4.1.2.

2.3 Inclusivity and the first person complex

Inclusivity is associated with the non-singular dimension but is not in itself part of the dimension of number. It relates to some of the possible interpretations of non-singular first persons, namely those that do not involve simply more than one speaker as represented in (15b, c, d) and illustrated on the basis of examples from English in (16b, c, d).

- (15) 1+1 more than one speaker
 1+2 the speaker and addressee
 1+2+3 the speaker, addressee and minimally one other
 1+3 the speaker and other

- (16) a. We solemnly swear
 b. We've got a bond in common, you and I.
 c. You, Anne and I are working ourselves to death.
 d. Me and Sarah Jones, we went up early.

Although none of the interpretations of the first person non-singular specified in (15) receive special encoding in English, all but the first (1+1) do so on a cross-linguistic basis. The encoding possibilities of the other three interpretations that have been attested cross-linguistically are summarized in (17), taken from Siewierska and Bakker (2005).⁹

(17) NoWe		
Unified We	1+2 = 1+2+3 = 1+3	
Only Inclusive	1+2 = 1+2+3	
Minimal Inclusive	1+2 vs. 1+2+3 = 1+3	
Augmented Inclusive	1+2+3 vs. 1+2 = 1+3	
Inclusive–exclusive	1+2 = 1+2+3 vs. 1+3	
Minimal Augmented	1+2 vs. 1+2+3 vs. 1+3	

The label ‘NoWe’ is used for paradigms which have no separate form for non-singular first person, i.e. they have no ‘we’ as distinct from ‘I’. The label ‘unified we’ (abbreviated UnifWe) embraces paradigms which have one form for non-singular first person covering all three interpretations, 1+2, 1+2+3 and 1+3, as is the case in English and most of the person paradigms cited earlier. We will consider this form to belong to plural number in case there is such a paradigm. The third pattern labelled ‘only inclusive’ (OnlyIncl) is used for paradigms in which there is a special form for 1+2 and 1+2+3 but not for the exclusive 1+3. An example of such a paradigm is given in (18) from Chalcatongo Mixtec, a language spoken in south-central Mexico, in which this pattern occurs both in the independent person forms and the corresponding SA clitics.

(18) Chalcatongo Mixtec (Macaulay 1996: 139)			
1	rú?ú	1+2	žó?ó
2	ró?ó	1+2+3	žó?ó
3M	càà		
3F	nā?ã		
3ANIMAL	kiti		
3SUPERNATURAL	í?a, íža		

The fourth pattern called ‘minimal inclusive’ (MinIncl) denotes paradigms such as the one in (19) from the Austronesian language Uma spoken in Sulawesi in which there is a separate form for the speaker–hearer dyad 1+2 and another form covering both 1+2+3 and 1+3.¹⁰

(19) Uma (Martens 1988: 169)

	Singular		Non-singular
1	aku'	1+2	kita'
		1+2+3	kai'
		1+3	kai'
2	iko	2+2	koi
3	hi'a	3+3	hira'

In the next pattern, called augmented inclusive (AugIncl), there is a special form for 1+2+3 and another form covering 1+2 and 1+3, as we see in the paradigm in (20) from Bunuba, a language of Western Australia.

(20) Bunuba (Rumsey 1999: 138)

	Singular		Non-singular
1	ngayini	1+2	ngiyirri
		1+2+3	yaarri
		1+3	ngiyirri
2	nginji	2+2	yinggirri
3	niy	3+3	biyirri

The penultimate pattern is the traditional inclusive–exclusive distinction (Incl–Excl) which involves the separation of groups involving the speaker and addressee (1+2 and 1+2+3) from those that involve the speaker and some other party (1+3). This pattern can be seen in several of the paradigms given earlier, for example, in (2) from Sye. And finally there is the three way split within the first person non-singular which we call the minimal augmented (MinAug) where there are separate forms for the minimal inclusive, augmented inclusive and exclusive. This pattern is illustrated in (21) from Koh, a language of Cameroon.

(21) Koh (Glidden 1985: 230)

	Singular		Non-singular
1	mì	1+2	ná
		1+2+3	nári
		1+3	bburu
2	mù	ì	
3	ka	i	

The above encodings may be grouped in terms of the number of formal distinctions made within the first person complex, as in (22) or the degree of separation of the speaker–addressee dyad, as in (23).

(22) MinAug > In/Excl > OnlyIncl > NoWe
 AugIncl UnifWe
 MinIncl

(23) MinAug > In/Excl > AugIncl > OnlyIncl > UnifWe > NoWe
 MinIncl

In assessing the degree of elaboration of the first person complex we have used the latter scale, with adaptations for higher numbers. This has resulted in a scale from 8 to 0 points. The MinAug has been assigned 8 points, the Incl-Excl 4 points, the AugIncl or MinIncl 3 points, the Only Incl 2 points, the UnifWe 1 point and NoWe 0 points. Paradigms with the Incl-Excl contrast in more than in one number opposition have received an extra point for each number opposition in which the Incl-Excl contrast obtains. Some minor patterns, for instance, the existence of an Incl-Excl contrast just in the plural or dual or alternatively a dual just in the inclusive or the exclusive have been treated as split and assigned 6 points.

2.4 Gender

Unlike in the case of number, which tends to be distributed uniformly across each of the three persons, gender strongly favors the third person. Gender in the second person is rather uncommon outside of North Africa, where it occurs in the Semitic, Berber, Cushitic and Chadic languages. And gender in the first person is quite rare (see e.g. Siewierska 2004: 105). A paradigm which quite exceptionally manifests gender in all three persons is given in (24) from Ngala, a Papuan language of the Sko family.

(24) Ngala (Laycok 1965: 133)

	SG	DU	PL
1M	wn	1ʌyn	1nan
1F	ñʊn		
2M	mʊn	2bʊn	2gwn
2F	yn		
3M	kʊr	3(kʊ)bʊr	3rʌr
3F	yn		

The likelihood of a particular person displaying gender is captured in the hierarchy in (25).

(25) 3 > 2 > 1

Gender is not only typical of the third person as opposed to the second or first but also of the singular rather than the non-singular. Moreover, among the non-singular

categories it seems to favor the more restricted ones, i.e. the dual or dual and trial as opposed to the plural, at least in the case of sex-based gender. Since gender in other than the third person is uncommon and is also typically restricted to the singular, we have not treated the absence of gender say in the second person as opposed to the third or the non-singular as opposed to the singular as involving homophony. Rather, in assessing the degree of elaboration of a paradigm for gender we have focused on the person–number combinations that do display a gender distinction. In assessing gender relative to person we have assigned 2 points for gender in the first or second person and 1 point for gender in the third. Additional points have been given for the presence of gender in other than the singular. The maximal number of points assigned in the languages in our sample is 5.

2.5 Summary

Using the point scales outlined above, we have assigned the person paradigms of free forms and bound S forms occurring in each language a set of four numerical values which together define a matrix such as the one in Table 1 which provides the values for the two person paradigms in Sye shown earlier in (2).

Table 1. Values for each semantic dimension per person form

Person form	Person Max = 4	Number Max = 6	Inclusivity Max = 8	Gender Max = 5
Free	4	2	4	0
Bound	4	4	5	0

The values on each of the semantic dimensions have then served as the basis of our comparison of the degree of semantic elaboration of free and bound forms within a language.

3. Free vs. bound pronoun languages and the head vs. dependent marking typology

The classification of languages in terms of whether the free or the bound person forms are the primary exponents of the grammatical category person into free pronoun vs. bound pronoun languages is highly reminiscent of the head vs. dependent marking typology elaborated by Nichols (1992). In the latter typology the major contributor to a positive value for head–marking is the availability of bound person forms for argument functions. The major contributor to a positive value for dependent marking, on the other hand, is the presence of case marking on free form

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

arguments. While the availability of bound person forms for argument functions does not logically entail that such forms must be viewed as the primary exponents of the grammatical category person, cross-linguistic studies suggest that this is indeed so. In fact under some analyses the bound person forms in head marking languages and not the free forms are considered to be the realizations of the verbal arguments. In any case free person forms in such languages are seen to be used only sparingly, typically in restricted discourse contexts. Accordingly, one would expect there to be a high degree of correlation between head marking languages and bound pronoun ones. An analogous correlation, though perhaps somewhat weaker, is also to be expected between dependent marking languages and free pronoun ones. Again although overt case marking of free form arguments does not logically entail the unavailability of bound person forms or the availability of such forms just for one of the argument functions, cross-linguistic investigations reveal that this is the dominant pattern. And needless to say, if bound forms are available only for one argument function, typically the subject, it is the free forms rather than the bound which will emerge as the primary exponents of person.

Another significant point of similarity between the two typologies is that both are scalar rather than discrete. Thus languages may display various degrees of head and dependent marking and differ with respect to the availability and use of free and bound person forms. In relation to the second point, it is well known that in languages which do have bound person forms these forms need not be obligatory. Thus, for example, in Makuchi, Pari and Retuara the bound person forms for both the subject and object are in complementary distribution with the free. The same applies with respect to the subject forms in, for instance, Coptic, and Teribe, and the object forms in Candoshi, Guarani, Kera, Kiribatese, Noon and Waura. What is less well known is that free forms are not necessarily available for all grammatical functions. For example, in Kiribatese, an Austronesian language spoken in the Republic of Kiribati, free person forms are used only as single word answers to questions (26a), and as subjects of nominal predicates (26b).

(26) Kiribatese (Groves et al. 1985: 64, 104)

- a. Antalae e roko? Ngala.
 who 3SG come (s)he
 ‘Who came? (S)he did.’
- b. Ngala te teretitenti.
 (s)he the president
 ‘(S)he is the president.’

As illustrated below, subjects of adjectival (27a) and verbal predicates (27bc) are indicated by weak person forms, and direct objects (27b) and objects of prepositions (27c) by person suffixes.

(27) Kiribatese (Groves et al. 1985: 106, 86, 110)

- a. E tikiraoi.
3SG pretty
'(S)he is pretty.'
- b. E noora-i.
3SG see-1SG
'He saw me.'
- c. E kanakoa te reta nako-iu
3SG sent the letter to-1SG
'(S)he sent the letter to me.'

Significantly, there are no free person forms for direct objects. Other languages which have free forms for at least some types of subjects but not for objects are Anejom, Au, Canela Kraho, Gapun, Geez, Malak Malak, Maranguku, Palikur, Salinan and Sumerian.

While both the free vs. bound pronoun typology and the head vs. dependent marking one are matters of degree, only the latter has actually been quantified. In view of this, we have opted to test the hypotheses relating to the distribution of semantic distinctions in free as opposed to bound person forms not directly relative to the free vs. bound pronoun language typology but relative to the head vs. dependent marking one.

Our classification of languages in terms of the head vs. dependent marking typology is based on the algorithm adopted by Nichols (1992), which involves adding up all the head marking and depending marking points assigned to clausal arguments on the basis of the marking characteristics that they display and dividing the dependent marking points by the sum of the dependent and head marking ones. Following Nichols, we have classified languages as head-marking if they score between 0.0 and 0.3 points, as mixed if they score between 0.4 and 0.6 points and as dependent-marking if they score between 0.7 and 1.00 point. In terms of this point scale, of the 328 languages in the sample 109 (33%) emerge as head marking ones, 145 (44%) as mixed, and 74 (23%) as dependent marking languages.¹¹

4. Semantic distinctions in person forms and head vs. dependent marking

The hypotheses under investigation lead us to expect that there should be a difference between head marking and dependent marking languages with respect to the encoding of semantic distinctions in free and bound person forms. The difference in question is, however, seen to obtain only in the case of languages in which the semantic distinctions expressed in free and bound forms are not the same. Therefore, in or-

der to have some idea of the actual scope of the phenomenon to be investigated, let us first consider the extent of same vs. different encoding of person, number, inclusivity and gender in free and bound person forms among the languages in the sample.

In comparing free and bound person forms with respect to the semantic distinctions that they display we have taken two paradigms as displaying the same semantic distinctions with respect to a given semantic dimension only if they achieve identical scores on the numerical scales for the dimension in question. The differences have been divided into two groups, those in which the semantic distinctions displayed by free forms are more elaborate than those manifested by the bound forms ($F > B$), and the converse, i.e. those in which the semantic distinctions displayed by the bound forms are more elaborate than those present in the free forms ($F < B$). In relation to the dimension of person we have performed two types of calculations. The first does not consider as third person forms, forms which are homophonous with demonstratives, while the second does. As we shall see below, the above difference in the treatment of third person forms homophonous with demonstratives has considerable consequences on the number of instances of $F < B$ in the sample. The distribution of the same as opposed to different semantic distinctions and the two types of differences along each of the four semantic dimensions is shown in Table 2a and Table 2b. In Table 2a the third person forms homophonous with demonstratives have been disregarded; in Table 2b they have been included.

We see that in the vast majority of instances the degree of elaboration of each of the four semantic dimensions in free and bound person forms is the same.¹² The largest percentage of none quality of 38.5% is with respect to person in Table 2a, i.e. when demonstratives are disregarded. The differences along the other dimensions and also of person when demonstratives are included are all well under 30%. The hypotheses to be tested thus pertain only to the 347 (Table 2a) and 307 (Table 2b) cases manifesting a difference along any of the four semantic dimensions.¹³

Table 2a. Differences between Free forms and bound S forms N = 328 3#Dem

Comparison	Person	Number	Gender	Inclusivity
F = B N = 965	202 (61.5%)	239 (72.82%)	265 (80.8%)	259 (79.0%)
F > B N = 239	68 (20.7%)	65 (19.8%)	50 (15.2%)	56 (17.0%)
F < B N = 108	58 (17.6%)	24 (7.3%)	13 (4.0%)	13 (4.0%)

Table 2b. Differences between Free forms and bound S forms N = 328

Comparison	Person	Number	Gender	Inclusivity
F = B N = 1005	242 (73.8%)	239 (72.82%)	265 (80.8%)	259 (79.0%)
F > B N = 253	82 (25.0%)	65 (19.8%)	50 (15.2%)	56 (17.0%)
F < B N = 54	4 (1.2%)	24 (7.3%)	13 (4.0%)	13 (4.0%)

The figures in Table 2a and 2b reveal a significant contrast between the instances of $F > B$ as compared to $F < B$ on every semantic dimension, especially person. This is fully in line with the grammaticalization scenario which leads us to expect that bound person forms should be equally or less well differentiated semantically than their free counterparts. The contrast with respect to the instances of $F > B$ as opposed to $F < B$ is particularly striking in the case of the figures in Table 2b, in which the overall instances of $F > B$ outnumber those of $F < B$ by nearly 5 to 1. However, the corresponding figure for the data in Table 2a is only 2 to 1. This difference in the proportion of $F > B$ to $F < B$ in the two Tables has quite divergent implications for the hypotheses that we are seeking to examine. The very low number of instances of $F < B$ as compared to $F > B$ in Table 2b casts doubt on the validity of any hypothesis positing an asymmetry with respect to the degree of elaboration of semantic distinctions in free and bound person forms which is bi-directional rather than uni-directional. But no such doubts arise with respect to the data in Table 2a where the levels of $F > B$ as opposed to $F < B$ are much more comparable. In all, though the data reveal a clear preference for $F > B$ over $F < B$, some type of bi-directional asymmetry in the degree of elaboration of the encoding of semantic distinctions in free and bound person forms may well emerge once the data are considered relative to the head vs. dependent marking typology. It is to this that we now turn.

4.1 Hypothesis 1

The first hypothesis is that if free and bound person forms differ in terms of the semantic distinctions that they encode, the semantic distinctions encoded in free forms should be more elaborate than those in bound forms in dependent marking languages, while the converse should be the case in head marking languages. The most liberal interpretation of this hypothesis is that: (a) there should be more instances of $F > B$ in dependent marking languages than in head marking languages and conversely (b) there should be more instances of $F < B$ in head marking than in dependent marking languages.¹⁴ For ease of reference we shall refer to these two sub-hypotheses as H1a and H1b respectively. Let us now consider how these two sub-hypotheses fare relative to each of the four semantic dimensions.

4.1.1 *Person*

In considering H1a and H1b in relation to the semantic dimension of person, number and gender we have taken into account both differences affecting any of the three persons and also those affecting only the speech act participants. Therefore in the relevant Tables there is more than one row of figures for both $F > B$ and $F < B$. In the case of person separate figures are also provided for cases in which demon-

Table 3. Differences in person distinctions in head, mixed and dependent marking languages

	H-marking languages	M-marking languages	D-marking languages
F > B ₁₂₃	24 (22.0%)	33 (22.6%)	25 (33.8%)
F > B _{123 3#Dem}	21 (19.2%)	25 (17.8%)	22 (29.7%)
F > B ₁₂	7 (6.4%)	12 (8.3%)	18 (24.3%)
F < B ₁₂₃	3 (2.8%)	1 (0.7%)	0 (0.0%)
F < B _{123 3#Dem}	18 (16.5%)	28 (19.9%)	12 (16.2%)
F < B ₁₂	0 (0.0%)	0 (0.0%)	0 (0.0%)

stratives functioning as third person forms have and have not been included. For person thus there are three rows of figures for F > B and F < B respectively and for number and gender two. Since inclusivity is tied to the first person complex, only one set of figures is provided.

The distribution of more elaborate person distinctions in free forms than in bound forms F > B, and of more elaborate person distinctions in bound forms than in free forms F < B respectively in the head, mixed and dependent marking languages in the sample is shown in Table 3.

We see that in relation to person H1a is borne out by the data. There are indeed more instances of F > B in dependent than in head marking languages irrespective of whether demonstratives are or are not treated as third person forms and irrespective of whether all persons or just the first and second are considered. The difference is particularly clear with respect to speech act participants, which evince more elaborate encoding in free forms as compared to bound forms nearly four times as often in dependent marking languages than in head marking languages. By contrast H1b fares not so well. It finds some support in the data when demonstratives are included. There are no dependent marking languages with F < B but there are such head and mixed marking languages. When demonstratives functioning as third person forms are excluded, however, the percentage of F < B in head marking and dependent marking languages is virtually the same and counter to what one would expect, the highest level of F < B occurs in mixed marking languages. Needless to say, all the cases of F < B involve the third person, as illustrated in (28) on the basis of Acoma, a language of New Mexico.

Table 4. Differences in number distinctions in head, mixed and dependent marking languages

	H-marking languages	M-marking languages	D-marking languages
F > B ₁₂₃	24 (22.0%)	22 (15.1%)	19 (25.6%)
F > B ₁₂	10 (9.1%)	6 (4.1%)	2 (2.7%)
F < B ₁₂₃	6 (5.5%)	15 (10.3%)	3 (4.1%)
F < B ₁₂	2 (1.8%)	5 (3.4%)	2 (2.7%)

(28) Acoma (Miller 1965: 174, 100)

	Free	S prefixes
1SG	sinume, hinume	1 s-
2SG	hisume	2 ş-
		3 g-/s-

4.1.2 *Number*

The relevant data pertaining to H1a and H1b involving the dimension of number are presented in Table 4.

The data are less supportive of either H1a or H1b than in the case of person. The predicted asymmetry between head and dependent marking languages is borne out when all three persons are considered. In line with H1a the percentage of dependent marking languages with $F > B$ is higher than of the head marking languages. And in accordance with H1b the percentage of head marking languages with $F < B$ is higher than of dependent marking languages. The mixed marking languages, however, do not behave as being in between head and dependent marking ones. The percentage of $F > B$ is lower than in the head-marking languages (rather than higher) and that of $F < B$ nearly twice as high as in the head marking languages (rather than lower). When number is viewed only in relation to the first and second person, there is no support for either the H1a or the H1b. The data are in fact consistent with predictions that are the very opposite to those captured by the H1a and H1b. The percentage of head marking languages with first and/or second persons manifesting richer number distinctions in the free forms than in the bound is three times higher than that of dependent marking languages. Conversely, the percentage of dependent marking languages with more elaboration for number in speech act participants in bound forms than in free forms is marginally higher than of head marking languages. Again, however, the mixed languages do not behave as might be expected.

4.1.3 *Inclusivity*

A yet different picture of H1a and H1b emerges when they are considered with respect to the dimension of inclusivity. As H1a predicts, in dependent marking languages the dimension of inclusivity is better elaborated in free than in bound person forms ($F > B$), in fact about 50 per cent more often than in head marking languages. Contrary to H1b, on the other hand, the highest level of $F < B$ with respect to inclusivity is manifested in mixed marking languages such as the West Papuan language Hatam, the person forms of which are illustrated in (29), rather than in head marking ones.

Table 5. Differences in inclusivity distinctions in head, mixed and dependent marking languages

	H-marking languages	M-marking languages	D-marking languages
F > B	16 (14.6%)	24 (16.6%)	16 (21.6%)
F < B	2 (1.8%)	8 (5.5%)	3 (4.1%)

(29) Hatam (Reesink 1999: 40)¹⁵

	Free		S
1	da	1	di-
2	na	2	a-
3	no(k)	3	Ø/ni-
1+2	sa	1+2	si-
1+2+3	nye	1+2+3	i(g)-
1+3	nye	1+3	<u>ni</u> -
2+2	je	2+2	ji-
3+3	yo(k)	3+3	i(g)-

Moreover, the level of F < B in head marking languages is twice as low as in the dependent marking languages. The relevant data are presented in Table 5.

4.1.4 Gender

For the semantic dimension of gender there is evidence for H1a and extremely weak evidence for H1b when all three persons are concerned. It is indeed the case that the percentage of dependent marking languages with finer distinctions in gender in the free forms as compared to the bound is higher than that of head marking languages. It is also the case that the percentage of head marking languages is minimally higher than of the dependent marking languages, in which the bound forms overall are richer in gender. However, with respect to just the first and second person neither H1a nor H1b holds. It is in fact the head marking rather than the dependent marking languages which exhibit relatively more gender elaboration in the free forms. And also counter to H1b, it is the dependent marking rather than the head marking languages which exhibit more gender distinctions in the bound forms. However, the latter is due to the presence of a gender distinction in the first person bound forms of only one dependent marking language, namely the Colombian language Cubeo. The bound markers in this language are primarily evidentiality and gender/number markers, which only secondarily mark person. As illustrated in (30), which presents the free forms and the bound forms used for witnessed events in the non-recent past and present habitual in the indicative mood, whereas the free forms manifest a gender contrast in the third person singular, the bound have a gender contrast both in the first and third persons singular.

(30) Cubeo (Morse and Maxwell 1999: 42, 80)

	Free	S
1SG	jì	1SG M -kaki
2SG	bì	1SG F -kako
3SG M	ì	2SG -Aw
3SG F	õ	3SG M -Abē
1INCL	bāxā	3SG F -Ako
1EXCL	jì xā	3INAN -Awì
2PL	bì xā	1INCL -Awì
3PL	dā	1EXCL -karā
		2PL -Awì
		3PL -Ibā

Note that the first person inclusive, the second person and the third inanimate are all homophonous. The data reflecting the distribution of gender in the head, mixed and dependent marking languages in the sample described above are presented in Table 6.

4.1.5 Summary

The extent to which H1a and H1b are borne out by the relevant languages in our sample for each of the four semantic dimensions is summarised in Table 7.

We see that the four semantic dimensions differ with respect to the degree of support that they offer for the two sub-hypotheses. The strongest degree of support comes from person with respect for H1a. H1b is weakly supported and only when demonstratives are included. Inclusivity provides support for H1a but not for H1b. The evidence from number and gender is inconclusive in that the predicted

Table 6. Differences in gender distinctions in head, mixed and dependent marking languages

	H-marking languages	M-marking languages	D-marking languages
F > B ₁₂₃	15 (13.8%)	21 (14.5%)	14 (18.9%)
F > B ₁₂	3 (2.7%)	5 (3.4%)	1 (1.4%)
F < B ₁₂₃	6 (5.5%)	3 (2.1%)	4 (5.4%)
F < B ₁₂	0 (0.0%)	0 (0.0%)	1 (1.4%)

Table 7. The correctness of H1a and H1b among the relevant languages in the sample

	Person		Number		Inclusivity	Gender	
	All	1st & 2nd	All	1st & 2nd		All	1st & 2nd
H1a	+	+	(+)	-	+	+	-
H1b	+	?	(+)	-	-	+	-

asymmetry between head and dependent marking languages is displayed when all persons are considered but not with respect to the speech act participants. Moreover in the case of number, the mixed marking languages behave in an unexpected way. The fact that four semantic dimensions do not behave identically and that for number and gender the hypotheses fare quite differently depending on which persons are taken into account suggests that it is worth considering hypothesis 2.

4.2 Hypothesis 2

The second hypothesis makes different predictions with respect to the distribution of person and inclusivity as opposed to number and gender. What unites person and inclusivity is that they are inherently tied to the discourse roles of speaker and addressee. Number and gender by contrast have to do with the nature of the referents filling the discourse roles of speaker and addressee. The basic assumption underlying hypothesis 2 is that the primary exponents of the category person should not feature information which would detract from their ability to fulfil their primary function which is the identification of the speaker and addressee. In other words, the primary exponents of the category person should be well differentiated with respect to person and inclusivity but should not evince elaborate gender or number distinctions. Thus given that the primary exponents of the grammatical category person in dependent marking languages are free forms and in head marking languages bound forms, hypothesis 2 is as follows:

If free and bound person forms differ in terms of the semantic distinctions that they encode, with respect to person and inclusivity free forms should be more elaborate than bound forms in dependent marking languages, and bound forms should be more elaborate than free forms in head marking languages, but with respect to number and gender, bound forms should be more elaborate than free forms in dependent marking languages, and free forms should be more elaborate than bound in head marking languages.

If we apply the same liberal interpretation of H2 as of H1, then we may expect (a) there to be more instances of $F_p > B_p$ and/or $F_i > B_i$ in dependent marking languages than in head marking language, (b) more instances of $F_p < B_p$ and/or $F_i < B_i$ in head marking languages than in dependent marking languages, (c) more instances of $F_n < B_n$ or $F_g < B_g$ in dependent marking languages than in head marking languages and (d) more instances of $F_n > B_n$ and/or $F_g > B_g$ in head marking than in dependent marking languages. We will refer to each of the above as H2a, H2b, H2c and H2d respectively.

The first two sub-hypotheses H2a and H2b are equivalent to H1a and H1b as applied to person and inclusivity. H2c and H2d, on the other hand, make the opposite

predictions to H1a and H1b. As evidenced by the data in Tables 3 and 5 discussed earlier, H2a is confirmed with respect to person and inclusivity. H2b, by contrast, is not confirmed. There are head marking languages in the sample in which the dimensions of person or inclusivity are more elaborate in the bound forms than in the free, but this is just as common or even more common in dependent marking and/or mixed marking languages. It is also of interest to note that the head marking languages in the sample which do confirm to H2b have free forms of the NoWe type and lacking number oppositions as is the case in the Tanoan language Kiowa, the person forms of which are illustrated in (31).

(31) Kiowa (Watkins 1984: 100, 113, 115)

	Free		S		
1	nó	1	à-	1+2+(3)	bà-
2	ám	2	èm-	1+3	è-
3	Ø-	2+2(+3)	bà-		
		Inverse	è-	3+3(hum)	á-

As for H2c and H2d, both find support along the dimensions of number and gender though with respect to speech act participants rather than for all persons. It is, however, precisely with respect to speech act participants that the distribution of number and gender is claimed to differ from that of person and inclusivity. As predicted by H2c, the percentage of dependent marking languages in which the first and second person bound forms have more elaborate number or gender distinctions than the free forms is higher than in head marking languages (2.7% vs. 1.8% and 1.4% vs. 0% respectively) and the percentage of head marking languages in which the free forms of the first and second person evince more number or gender distinctions than the bound forms is higher than in dependent marking languages (9.1% vs. 2.7% and 2.7% vs. 1.8% respectively). It must be remembered, however, that gender marking in speech act participants is in itself very rare. Accordingly, the number of cases supporting H2c and H2d with respect to gender is very low. Nonetheless, the data cited above do provide support for hypothesis 2. Moreover, the support is stronger than with respect to hypothesis 1. Now what remains to be seen is how these hypotheses can be reconciled with the grammaticalization scenario.

5. Bi-directional asymmetries in semantic distinctions and grammaticalization

We have seen that for every semantic dimension instances of $F > B$ are more common than of $F < B$ in all three types of languages, head marking, mixed marking and dependent marking.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

Thus contrary to hypothesis 1 and hypothesis 2 and in line with the grammaticalization scenario, there is no evident bi-directionality with respect to the asymmetries in free and bound person forms with respect to the semantic distinctions that they encode. Nonetheless we have also seen that there is some support for hypothesis 2, which predicts that the asymmetry in the semantic elaboration of person forms is to some extent sensitive to the nature of the semantic distinctions and type of languages involved. The support in question, however, comes in the main from the distribution of $F > B$ patterns not of $F < B$ ones. It therefore poses no threat to the grammaticalization scenario since $F > B$ patterns are precisely what this scenario predicts.

The above notwithstanding, it is of interest to consider what implications our findings do carry for the grammaticalization account of the asymmetries in semantic distinctions between free and bound forms. Our data have shown that there are some differences in the levels of $F > B$ between head, mixed and dependent marking languages. Dependent marking languages display higher levels of $F > B$ on each of the four semantic dimensions than the mixed and head marking ones. Within the grammaticalization context this may be interpreted as suggestive of the bound forms in dependent marking languages being diachronically older, more grammaticalized, than those in mixed and head marking languages. The only exceptions to the generally higher level of $F > B$ in dependent marking languages in our sample involve speech act participants along the dimensions of number and gender. In both instances the head marking languages surpass the dependent with respect to the level of $F > B$. It is this that constitutes the strongest support for hypothesis 2. From the perspective of grammaticalization the higher levels of $F > B$ in head marking languages in the case of speech act participants on the dimensions of number and gender are somewhat unexpected, since there is no obvious reason why the effects of grammaticalization which are typically stronger in dependent marking languages should in this case favour the head marking ones. On closer inspection, however, an alternative explanation for the data is available. Although the most obvious way of viewing the $F > B$ patterns of speech act participant with respect to number and gender in head marking languages is as a consequence of semantic bleaching on the part of the bound forms, it is also possible to see them as the result of the emergence of new free forms with more elaborate semantic distinctions subsequent to the development of the bound. In this case the lower degree of elaboration of number and gender in the bound forms as compared to the free would not be directly attributable to their relative diachronic age or grammaticalization. Currently we do not have enough data at our disposal to verify whether all the relevant instances of $F > B$ in the head marking languages in the sample could be thus analysed. However, supportive of such an analysis is the fact that in head marking languages the free forms often consist of the bound with an invariable stem such as the word for body or human or an emphatic particle or

the verb “to be”. Such free forms are thus diachronically younger than the bound.

Potentially more challenging for the grammaticalization scenario are the patterns of $F < B$. These, however, can be accommodated within the context of grammaticalization if the free forms are taken to have lost the relevant semantic distinctions after the emergence of the bound forms or if new free forms featuring less elaborate semantic distinctions are taken to have arisen subsequent to the development of the bound forms. Yet another option is the emergence of new semantic distinctions in the bound person forms.¹⁶ Whether there is any correspondence between these three possibilities and the four semantic dimensions and/or language type is not clear. Hypothesis 2 leads us to expect loss of semantic distinctions in free forms to be the source of $F < B$ patterns in dependent marking languages with respect to number and gender and the development of new less differentiated free forms to be the source of $F < B$ patterns in head marking languages in the case of person and inclusivity. Our investigation, however, has provided little evidence for the validity of the $F < B$ predictions of hypothesis 2. Crucially, the most numerous group of instances of $F < B$ in the sample which involves absence of third person free forms (when demonstratives are disregarded) does not favour any of the three marking types. Nor does it clearly favour any of the three sources of $F < B$ patterns listed above. Nonetheless it needs to be mentioned that many of the relevant instances of $F < B$ with respect to person involve third person bound forms which are covert. The bound paradigms qualify as exhibiting a three way person distinction rather than a two way one since the zero forms function as paradigmatic zeroes. Such cases of $F < B$ may be viewed as instantiations, though be it perhaps atypical ones, of the third source of $F < B$ patterns, i.e. of a new semantic distinction arising in bound forms.

6. Concluding remarks

Our investigation of the bi-directional as opposed to unidirectional asymmetry in the encoding of semantic distinctions in free and bound person forms has lent little support to the two bi-directionality hypotheses. Nonetheless it needs to be emphasised that for reasons of expediency our investigation has been carried out using a typology different from the one with reference to which the two hypotheses have been actually framed. Although as argued in Section 3 one may expect there to be a high degree of overlap between the head vs. dependent marking typology which we have employed here and the free pronoun vs. bound pronoun language typology which yet remains to be elaborated fully, overlap is not the same as identity. And given the relatively low number of instances of $F < B$, even minor differences in how languages are classified could have significant repercussions on the overall validity of the hypotheses. Our results therefore can only be viewed as suggestive but not conclusive.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

Abbreviations

ABS	absolute	NHUM	non-human
DEM	demonstrative	PAUC	paucal number
DU	dual number	PL	plural number
EXCL	exclusive	PRES	present tense
F	feminine gender	S	sole argument of intransitive clause
HUM	human	SG	singular number
INCL	inclusive	1	first person
M	masculine gender	2	second person
N	neuter gender	3	third person

Notes

1. The assumption that in some languages the bound person forms are the primary exponents of the category person is implicit in the so called Pronominal Argument Hypothesis developed by Jelinek (1984) for non-configurational languages within the context of Chomskian generative grammar and in the grammatical vs. pronominal agreement distinction elaborated by Bresnan and Mchombo (1987). In fact this analysis has been advocated for various subsets of languages by many other scholars ever since its first attested formulation by Wilhelm von Humboldt in the 1830s.
2. This second hypothesis has been advanced by Bhat (2004: 18–25).
3. For some criticism of the grammaticalization approach and in particular the relationship between semantic bleaching and phonological reduction see Campbell (2001) or Newmeyer (2001).
4. The term weak form is used in the sense of Bresnan (2001), i.e. not for just unstressed versions of free forms but rather for forms differing from free forms both phonologically and in syntactic distribution.
5. In languages which have more than one set of S forms used with unergative predicates, we have used those occurring in main, declarative clauses, in the main, in the past tense or realis mood.
6. We have considered as phonologically distinct also zero forms, provided that they function as paradigmatic zeroes, i.e. they are functionally equivalent to overt forms. This is typically the case with respect to missing bound forms, though not missing free forms.
7. This statement holds provided that in languages in which demonstratives are used in lieu of third person forms the demonstratives are treated as person forms. In investigating the two hypotheses we have used two analyses, one where demonstratives functioning as third person forms are included and one in which they are excluded.
8. Corbett (2000: 21–38) cautions that most trials are in fact paucals and therefore some paradigms traditionally interpreted as having both a trial and paucal or in the case of Sursurunga, a trial and a quadral are better interpreted as displaying a greater and lesser paucal.
9. The typology of the first person complex in (36) is a slight modification of the typology developed by Cysouw (2003).
10. Paradigms with a minimal inclusive pattern are typically treated as manifesting a dual confined to the first person. We, however, recognise the existence of a dual only if it is used in all person categories.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

11. The low level of dependent marking languages in the sample is due to the fact that given the nature of the investigation the only dependent marking languages that we selected were those which have some type of bound person forms for the S.
12. It must be mentioned that unlike in the case of person and number, no difference in gender or inclusivity between a paradigm of free forms and bound forms in the vast majority of cases involves absence of gender or inclusivity respectively.
13. The 307 differences in the encoding of semantic distinctions in free and bound person forms are distributed over xxx languages.
14. A strict interpretation of the first hypothesis would entail that in head marking languages there should be more instances of $F < B$ than of $F > B$ for every semantic dimension and in dependent marking languages there should be more instances of $F > B$ than of $F < B$ for every semantic dimension.
15. The free forms in Hatam display a minimal inclusive pattern, the bound forms a minimal augmented one.
16. An instance of such a development is discussed by Sakel (2003) in the Mosestenan family of Bolivia. A less typical instance of this source will also be mentioned below.

References

- Bhat, Darbhe Narayana Shankara. 2004. *Pronouns: A cross-linguistic study*. Oxford: Oxford University Press [Oxford Studies in Typology and Linguistic Theory].
- Bresnan, Joan. 2001. The emergence of the unmarked pronoun. *Optimality-theoretic syntax*, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 113–42. Cambridge, MA: MIT Press.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun and agreement in Chichewa. *Language* 63: 741–82.
- Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences* 23: 113–61.
- Casad, Eugene H. 1984. Cora. *Studies in Uto-Aztecan grammar*, vol. 4., ed. by Ronald W. Langacker, 151–459. Arlington, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.
- Cowan, Hendrik K. J. 1965. *Grammar of the Sentani language*. 's-Gravenhage: Martinus Nijhoff.
- Croft, William. 2000. *Explaining language change*. London: Longman.
- Crowley, Terry. 1998. *An Erromangan (Sye) grammar*. Honolulu: University of Hawai'i Press.
- Crowley, Terry. 2004. Southeastern Ambrym. *The Oceanic languages*, ed. by John Lynch, Malcolm Ross and Terry Crowley, 660–7. London: Curzon Press.
- Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press [Oxford Studies in Typology and Linguistic Theory].
- De Vries, Lourens. 1989. *Studies in Wambon and Kombai*. Ph.D. dissertation, University of Amsterdam.
- Dixon, Robert M. W. and Alexandra Y. Aikhenvald (eds.). 2002. *Word. A cross-linguistic typology*. Cambridge: Cambridge University Press.
- Dutton, Tom E. 1996. *Koiari*. Munich: Lincom Europa.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

- Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. *Subject and topic*, ed. by Charles N. Li, 151–88. New York: Academic Press.
- Glidden, Suellen H. 1985. The Koh verbal system. *Work papers of the Summer Institute of Linguistics, University of North Dakota* 29: 223–82.
- Groves, Terabata R., Gordon W. Groves and Roderick Jacobs. 1985. *Kiribatese: An outline description*. Canberra: Australian National University [Pacific Linguistics, D-64].
- Guirardello, Raquel. 1999. A reference grammar of Trumai. Ph.D dissertation, Rice University, Houston, TX.
- Haspelmath, Martin. 1998. Does grammaticalization need reanalysis? *Studies in Language* 22(2): 315–51.
- Hollenbach, Barbara E. 1992. A syntactic sketch of Copala Trique. *Studies in the syntax of Mixtecan Languages*, vol. 4., ed. by C. Henry Bradley and Barbara E. Hollenbrach, 173–431. Arlington, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Hyman, Larry M. 1981. *Noni grammatical structure*. Los Angeles: The Department of Linguistics, University of Southern California [Southern California Occasional papers in Linguistics 9].
- Irwin, Barry S. 1974. *Salt Yui grammar*. Canberra: Australian National University [Pacific Linguistics, B-35].
- Jelinek, Eloise. 1984. Empty categories and non-configurational languages. *Natural Language and Linguistic Theory* 2: 39–76.
- Laidig, Wyn D. 1993. Insights from Larike possessive constructions. *Oceanic Linguistics* 32: 312–51.
- Laycock, Donald C. 1965. *The Ndu language family*. Canberra: Australian National University [Linguistic Circle of Canberra Publications].
- Lehmann, Christian. 1982a. Universal and typological aspects of agreement. *Apprehension. Das sprachliche Erfassen von Gegenständen*, vol. 2, ed. by Hans-Jakob Seiler and Franz J. Stachowiak, 201–67. Tübingen: Gunter Narr.
- Lehmann, Christian. 1982b. *Thoughts on grammaticalization: A programmatic sketch*, vol. 1. Köln: Institut für Sprachwissenschaft [Arbeiten des Kölner Universalien-Projekts 48].
- Macaulay, Monica. 1996. *A grammar of Chalcatongo Mixtec*. Berkeley: University of California Press [University of California Publications in Linguistics 127].
- Martens, Michael P. 1988. Notes on Uma verbs. *Papers in Western Austronesian Linguistics* 4: 167–237.
- Miller, Wick R. 1965. *Acoma grammar and texts*. Berkeley: University of California Press.
- Morse, Nancy L. and Michael B. Maxwell. 1999. *Cubeo grammar*. Arlington, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Newmeyer, Frederick J. 2001. Deconstructing grammaticalization. *Language Sciences* 23: 187–229.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago and London: The University of Chicago Press.
- Reesink, Ger P. 1999. *A grammar of Hatam*. Canberra: The Australian National University [Pacific Linguistics C-146].
- Romankevicius Costa, Raquel G. 1998. Aspects of ergativity in Marubo (Panoan). *Journal of Amazonian Linguistics* 1: 50–103.

- Rumsey, Alan. 1982. *An intra-sentence grammar of Ungarinjin*. Canberra: The Australian National University [Pacific Linguistics B-86].
- Rumsey, Alan. 1996. On some relationships among person, number, and mode in Bunuba. *Studies in Kimberley languages in honour of Howard Coate*, ed. by William McGregor, 139–48. Munich: Lincom Europa.
- Sakel, Jeanette. 2003. *A grammar of Mose'tén*. Ph.D. dissertation, University of Nijmegen.
- Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: why objects don't make it. *Folia Linguistica* 33(2):225–51.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Siewierska, Anna and Dik Bakker. 2005. Inclusive and exclusive in free and bound person forms. *Clusivity. Typology and case studies of the inclusive–exclusive distinction*, ed. by Elena Filmonova. Amsterdam: John Benjamins.
- Sohn, Ho-min. 1975. *Woleaian reference grammar*. Honolulu: University Press of Hawai'i.
- Watkins, Laurel J. 1984. *A grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Zwicky, Arnold. 1985. Clitics and particles. *Language* 61:283–305.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY