'We' rules

The impact of an inclusive/exclusive opposition on the paradigmatic structure of person marking[*]

Michael Cysouw
ZAS Berlin

1. Introduction

Pronominal elements in language are traditionally analysed on the basis of a division into three persons and two numbers. Other categories that are found regularly, like duals or an inclusive/exclusive opposition, are often seen as additions to this pronominal core. This perspective might seem natural from the point of view of the structure of European languages, but it is rather Eurocentric to impose this analysis on the structure of other languages in the world (Forchheimer 1953; Greenberg 1963; Ingram 1978; Mühlhäusler and Harré 1990). This paper reports on a cross-linguistic study into the paradigmatic structure of pronominal elements. Specifically, the marking of person will be investigated. Person markers, either independent pronouns or pronominal inflection, are not considered as individual items but as bound in a paradigm. The structure of the paradigm determines the precise value of the person markers. One of the goals of investigating the cross-linguistic variety of grammaticalised categories in paradigms of person is to check empirically which kinds of homophony are accidental, and which are systematic — the assumption being that accidental homophony will only sporadically turn up in the world's languages, but systematic ones will occur regularly. An important result will be to show that an inclusive/exclusive opposition in a paradigm of person is a typologically salient feature — not a secondary addition. If an inclusive/ exclusive opposition is present, many paradigmatic structures are significantly less likely to occur. I will explain these restrictions by introducing the notion 'pure person'. Paradigms with an inclusive/exclusive opposition qualify as pure

person paradigms. I will claim that pure person paradigms are specialised to mark person categories. They are the real prototypical examples of person marking, and, consequently, the inclusive/exclusive opposition is not a secondary subcategory of pronominal marking.

In Section 2, the definition of the concept 'paradigm of person' adopted in this study is explicated. Some restrictions on its scope are necessary to keep the investigation within workable limits. The sampling method is discussed (see also the appendix) and the frame for the cross-linguistic classification will be presented. In Section 3, some general typological generalisations are discussed. I argue that the difference between accidental and systematic homophony is a continuum, not a clear dichotomy. Also in this section, the first signs are found that there is something special about paradigms that have an inclusive/exclusive opposition: these paradigms show less 'exceptional' paradigmatic structures. Section 4 investigates some possible correlates of the inclusive/exclusive opposition. The characteristics of paradigmatic structure that correlate significantly with the inclusive/exclusive opposition form a hierarchy. This Explicitness Hierarchy is described in Section 5. In that section, it is also noted that this hierarchy correlates with the morphological status of the paradigm. Highly explicit paradigms are more often marked with independent pronouns; barely explicit paradigms are more often inflectionally marked. This is precisely as would be expected from the perspective of grammaticalisation. However, diachronic change does not follow a unidirectional path along the hierarchy. Therefore, a different explanation is needed to extend the grammaticalisation account. Conscious awareness of the referential value of independent words (as opposed to less awareness of inflectionally marked categories) is invoked. Awareness of referential value favours more explicit marking of person categories. Section 6 summarises the central results and presents implications for future investigations.

2.   Definitions and method

The domain of inquiry of this investigation is the category of person. Person is defined prototypically as the linguistic marking of the speech roles 'speaker' and 'addressee'. For English, this includes words like *I* and *you*. These words specifically refer to the speaker and addressee — they can not be used for any other reference. An important definitional aspect is that only such *specialised* marking of person is included in this study. Many linguistic items can be used to refer to

the principal actors of a speech act besides such specialised person marking. Language is an interactional device, fine-tuned to the conversational setting, and, consequently, all kinds of nouns can be used to refer to the speaker or addressee. For example, when speaking to his child, a father can refer to himself by using the noun 'daddy' as in *Daddy is going upstairs.* Or in a school-class setting, a teacher can refer to a pupil by using a proper name, as in *How would John deal with this problem?* For English, such usage is restricted to specific settings and registers, yet in many other languages the usage of nouns and proper names to refer to speaker or addressee is much more widespread, e.g. Thai (Cooke 1968). Nonetheless, such non-specialised person markers are excluded from this investigation. Cross-linguistically, the exclusion of these markers amounts to a rather strong restriction of the investigation. Yet, such a formal criterion is necessary to assure the comparability of linguistic marking through the very diverse structures that are found among the world's languages.

It is beyond the scope of this study to discuss the overall structure of whole languages. This inquiry focuses specifically on the structure of the paradigm. A paradigm is defined as a group of morphemes that is syntagmatically identical. The morphemes that belong to a paradigm have, so to speak, the same place within the overall structure of the language. A paradigm of person markers, then, is a paradigm that includes at least an opposition between speaker and addressee.[1] For English, the person paradigm as defined by the primary opposition *I* versus *you* also includes *he, she, it, we* and *they*.[2] The cases that are counted in the present paper are individual person paradigms. One language can have (and usually has) more than one such paradigm.

Paradigms that mark person can be morphologically independent as well as inflectionally bound. I do not place any morphological or functional restrictions on the paradigms included in the sample, since in many languages pronominal inflection is functionally comparable to the independent pronouns of other languages.[3] For some languages, pronominal inflection is probably best qualified as agreement; for others the independent pronouns are optional adjuncts. In between these extremes, there is a large variety of functional possibilities attested for person marking. It would be of great interest to make a cross-linguistic investigation of this continuum. The present investigation is only a first step towards that goal. While the method of sampling adopted here allows only for an investigation of paradigmatic structure, the resulting classification can guide any future investigation into the syntagmatical structure of person marking. As a first step in this direction, I will show that morphological status turns out to be a significant parameter of variation in paradigmatic structure (see Section 5).

The data presented in this article are counts based on an extensive investigation of the variability of the paradigmatic structure of person marking by the present author (Cysouw 2001). The objective of that investigation was to present a survey of the variation in structure of person paradigms. Some regions or families turn out to be more interesting than others, leading to some genetic skewing in the data. However, the sample of 265 paradigms of person from 234 languages for the present study is large and genetically varied enough to qualify as a typological sound database (see the appendix for further details). Due to space limitations, I cannot discuss in detail the actual data used in this study, but only present the generalisations. I should point out that it would be rather easy to expand the number of the most common paradigmatic structures. I decided to stop collecting examples of a particular paradigmatic structure when I had about 40 cases, as this was enough to establish that the structure was clearly common among the world's languages. This decision implies that the current database of paradigmatic structure is more variable than the actual linguistic variation among the world's languages. For the generalisations that will be presented in this article, this methodological decision possibly results in *less* strong correlations and trends than would have been found using a stricter procedure of genetic sampling. It is highly improbable that the sample used for this paper leads to more significant correlations than would characterize the actual world's linguistic diversity.

In this paper, only paradigms with singular and general non-singular (i.e. 'plural') forms are included. All paradigms that distinguish gender, honorifics or specialised number categories (like dual, trial etc.) are excluded. However, a few paradigms with gender in the third person singular are included because otherwise some unusual paradigmatic structures would have been missing. Contrary to these omissions, some instances of the category that is traditionally called 'first person inclusive dual' are included in the present domain. The first person inclusive dual is an ambiguous category (Greenberg 1988). In some paradigms it is clearly a dual form (these are excluded here), but in other paradigms it is a specific kind of person marking (these are included in this paper). The criterion that is used to discriminate between these two possibilities is the presence of other dual morphemes in the paradigm. If there are other dual person markers in the paradigm, then the first person inclusive dual is interpreted as a dual and excluded from the discussion here. In contrast, if there are no other dual morphemes in the paradigm, except for this first person inclusive dual, this morpheme is interpreted as a specific kind of person marking, called the 'minimal inclusive' (McKay 1978).[4] A minimal inclusive is a category that

includes the present speaker and addressee (i.e. inclusive), but nothing else (i.e. minimal). That duality is epiphenomenal to this category is shown by the fact that in almost all instances, this apparent dual is the only dual in the whole language (Plank 1996: 130–131). Incorporating all these restrictions, eight referentially distinct person categories remain as the subject of the present paper. Other person categories are theoretically feasible, yet are unattested in the world's languages (Cysouw 2001: 66–73).[5] These eight person categories are shown in Figure 1. The present investigation includes only paradigms that distinguish maximally eight different morphemes, one for each of these categories. The crux of this paper is paradigms that distinguish less than these eight person categories. The main question to be dealt with is which categories are combined into the reference of a single morpheme.

The eight person categories are shown here in the representational frame that will be used throughout this paper to display the structure of paradigms of person. The outline as presented in Figure 1 is chosen as a working compromise among various language-specific structures. The representational frame is kept identical throughout to enhance comparability, possibly at the expense of language-specific preferences.

Figure 1. The paradigmatic frame with the eight basic person categories.

3. Typological characteristics: Accidental vs. systematic homophony

Not all the paradigmatic structures which are theoretically possible are attested among the world's languages. In the present sample of 265 paradigms, 62 different paradigmatic structures are attested. These 62 different paradigms include each and every paradigmatic structure that I have been able to find.

This means, for example, that I have also included the perfective suffixes of Kunimaipa, as shown in Figure 2, although they show a different structure from the imperfective suffixes in the same language, and they are also different from the independent pronouns (Pence 1968: 110; Geary 1977: 26). The structure, as shown in Figure 2, is, as far as I know, unique among the world's languages.

	...-ho	1+2
		1+2+3
1 ...-ho	...-gi	1+3
2 ...-ngi		2+3
3 ...-ha		3+3

Figure 2. An uncommon paradigmatic structure: Kunimaipa perfective suffixes.

Many of the 62 paradigmatic structures are only attested in one, or in a very small number of cases. A restricted set of paradigmatic structures is more widespread. Typically, such common paradigmatic structures are found throughout the world's languages, independent of areal and genetic bonds. An example of a common paradigmatic structure is the paradigm of the present suffixes of Latin, as shown in Figure 3.

		1+2
	...-mus	1+2+3
1 ...-o		1+3
2 ...-s	...-tis	2+3
3 ...-t	...-unt	3+3

Figure 3. A common paradigmatic structure: Latin present suffixes.

There is no clear-cut division between the common and uncommon paradigmatic structures. Rather, a continuous cline from common to uncommon paradigmatic structures is found. The Latin-type paradigm is one of the most

common patterns world-wide, yet it is surely not the only common paradigmatic structure. The structures of the nine most frequent paradigmatic structures are shown in Figure 4. These nine paradigmatic structures account for almost 70% of the sample. They form a rather coherent group, which I will refer to as the 'common' paradigmatic structures. The kinds of homophony that are found in these paradigmatic structures are clearly systematic. The other 53 paradigmatic structures that I have found (and other curious examples that probably exist, but that I missed) are classified as 'uncommon'. However, this does not imply that all the special kinds of homophony in these paradigms are accidental. The distinction between common and uncommon is a rather ad hoc division of a continuum of frequency. The most frequent uncommon paradigm is found almost equally frequent as the least frequent common paradigm (Cysouw 2001: 158f.).
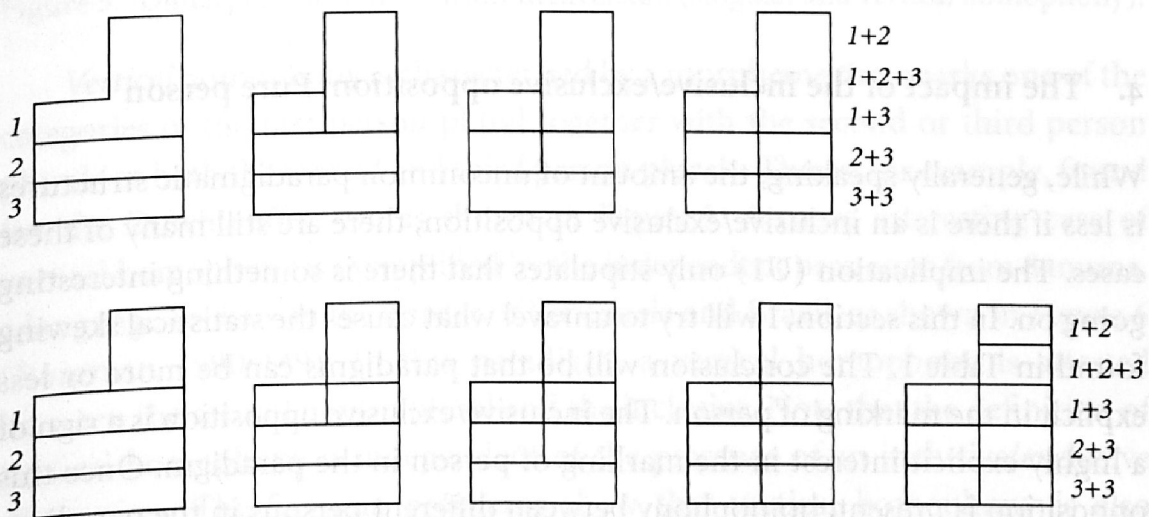

Figure 4. The nine most frequent paradigmatic structures.

A central characteristic of the structure of a pronominal paradigm is the way in which the categories of the first person plural are marked. The opposition between 1+2/1+2+3 (inclusive) and 1+3 (exclusive) turns out to be a good predictor for some other characteristics of the paradigmatic structure. A general typological observation is that the amount of paradigmatic variation is less for paradigms that have an inclusive/exclusive opposition, when compared to the paradigms that do not have this opposition. In other words, if there is an inclusive/exclusive opposition, then much fewer uncommon structures are attested. The frequencies from the sample are shown in Table 1. From these frequencies, a statistically valid implication can be extracted (Fisher's Exact $p = .001$):[6]

(U1)   Inclusive/exclusive opposition → less uncommon paradigmatic structures

Table 1. Paradigms with an inclusive/exclusive distinction show less uncommon paradigmatic structures

	Inclusive vs. Exclusive		
	No	Yes	Total
Common Paradigmatic Structures	87 (32.8%)	97 (36.6%)	184
Uncommon Paradigmatic Structures	57 (21.5%)	24 (9.1%)	81
Total	144	121	265

4.   The impact of the inclusive/exclusive opposition: Pure person

While, generally speaking, the amount of uncommon paradigmatic structures is less if there is an inclusive/exclusive opposition, there are still many of these cases. The implication (U1) only stipulates that there is something interesting going on. In this section, I will try to unravel what causes the statistical skewing found in Table 1. The conclusion will be that paradigms can be more or less explicit in the marking of person. The inclusive/exclusive opposition is a sign of a highly explicit interest in the marking of person in the paradigm. Once this opposition is present, homophony between different persons in the paradigm is rarely found, and, consequently, much fewer uncommon paradigmatic structures are attested.

Three different kinds of homophony, which I will call 'singular', 'vertical' and 'horizontal' homophony, are used to analyse the structure of a paradigm. The term homophony is used purely descriptively. Homophony simply means that some of the eight referential person categories (from Figure 1) are marked by the same morpheme. All kinds of homophony are treated alike — only after their distribution has been investigated there might be made a claim for a difference between accidental and systematic homophony.

Singular homophony is characterised by a homophony between two of the three singular categories. For example, in the Dutch present suffixes, the second and third person singular are marked identically, as shown in Figure 5. Other kinds of singular homophony that are attested are combinations of the first and

second person (as, for example, in the English present suffixes, marked by zero) and combinations of the first and third person (as, for example, in the Spanish 'pretérito imperfecto', marked by -*ba*).

			1+2
			1+2+3
1	...-Ø	...-en	1+3
2			2+3
	...-t		
3			3+3

Figure 5. Dutch present suffixes, without inversion (singular and vertical homophony).

Vertical homophony is characterised by a morpheme that marks one of the categories of the first person plural together with the second or third person plural (or both the second and third person plural). This is, for example, found for the Dutch suffix -*en*, as shown in Figure 5. A more interesting case of vertical homophony is exemplified by the independent pronouns from Sanuma, a language spoken on the border of Venezuela and Brazil, as shown in Figure 6 (Borgman 1990: 149). In this paradigm, a vertical homophony is attested between the second person plural and the inclusive. Note that the definition of vertical homophony is independent of the presence of an inclusive/exclusive distinction. The Sanuma pronouns show that vertical homophony is also possible with an inclusive/exclusive distinction.
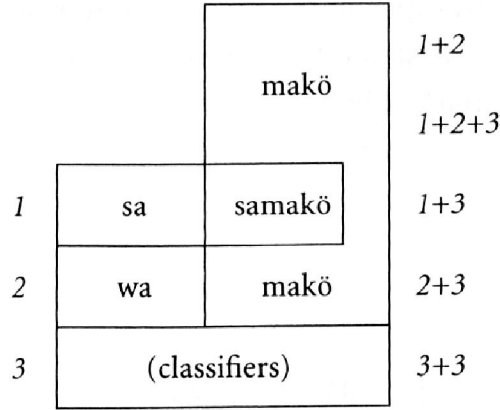
		makö	1+2
			1+2+3
1	sa	samakö	1+3
2	wa	makö	2+3
3	(classifiers)		3+3

Figure 6. Sanuma independent pronouns (vertical homophony).

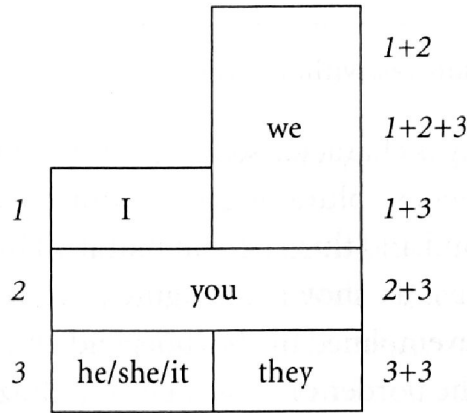|		we	1+2
			1+2+3
1	I		1+3
2	you		2+3
3	he/she/it	they	3+3

Figure 7. English subject pronouns (horizontal homophony).

Finally, *horizontal* homophony is characterised by a morpheme that marks both a singular and a non-singular category. This is exemplified here by the English independent pronouns, as shown in Figure 7. The pronoun *you* is used both for the singular and the plural address: a case of horizontal homophony. Horizontal homophony is a commonly occurring phenomenon, as can be seen from the structures shown in Figure 4 above. The vast majority of cases show a 'real' horizontal homophony between the first person singular and plural, the second person singular and plural or the third person singular and plural. Yet, there is a small set of cases that have a 'diagonal' homophony between singular and plural (seven cases in the sample, which is 2.6%). This is found, for example, in the German present suffixes, as shown in Figure 8 (see p. 52). In this paradigm, a homophony is found between the second person singular and the second person plural, marked by *-t*. Note also that a vertical homophony between the first and third person plural is found in this paradigm, marked by *-en*.

The presence or absence of these three kinds of homophony is now cross-cut with the occurrence of an inclusive/exclusive opposition within the same paradigm. First, the inclusive/exclusive opposition is inversely correlated with singular homophony, as shown in Table 2. As can be seen in the table, paradigms with an inclusive/exclusive opposition do not show any singular homophony at all. This results in a very strong implication (Fisher's exact $p = .000$):

(U2)   inclusive/exclusive opposition → no singular homophony

Table 2. Inclusive/exclusive opposition and singular homophony

	Inclusive vs. Exclusive		
	No	Yes	Total
No Singular Homophony	119 (44.9%)	121 (45.7%)	240
With Singular Homophony	25 (9.4%)	0 (0.0%)	25
Total	144	121	265

The inclusive/exclusive opposition is also inversely correlated with vertical homophony, as shown in Table 3. Paradigms with an inclusive/exclusive opposition seldom show vertical homophony. However, there are 14 (5.3%) exceptions to the following implication, which is still highly significant (Fisher's exact $p = .001$):

(U3)   inclusive/exclusive opposition → less vertical homophony

Table 3. Inclusive/exclusive opposition and vertical homophony

	Inclusive vs. Exclusive		
	No	Yes	Total
No Vertical Homophony	103 (38.9%)	107 (40.4%)	210
With Vertical Homophony	41 (15.5%)	14 (5.3%)	55
Total	144	121	265

The inclusive/exclusive opposition implies a low occurrence of vertical homophony. Formulated differently, the absence of vertical homophony is a precondition for the presence of an inclusive/exclusive opposition. The precondition

			1+2
		...-en	1+2+3
1	...-e		1+3
2	...-st	...-t	2+3
3	...-t	...-en	3+3

Figure 8. German present suffixes (diagonal and vertical homophony).

can be understood by considering vertical homophony as a kind of fused marking of the non-singular categories. This fusion should be absent for an inclusive/exclusive opposition to be feasible. Within the group of 14 exceptions to this generalisation, by far the most common variants of vertical homophony combine inclusive with second person plural (five cases) or exclusive with third person plural (four cases). These patterns can be referentially summarised by noting that the *inclusion of the addressee* is relevant for both the inclusive and the second person plural; oppositely, the *exclusion of the addressee* is relevant for both the exclusive and the third person plural. These exceptions are still neatly organised along the lines of the person categories. Yet, it is the addressee that is the crux of this marking instead of the speaker. The remaining five really exceptional cases are a collection of random kinds of homophony that can be characterised as accidental.

Finally, the inclusive/exclusive opposition is not at all correlated with horizontal homophony, as is shown in Table 4 (Fisher's exact $p = .388$). Statistically speaking, there is no reason to draw any conclusions from the slight skewing that is found in the table. The amount of skewing falls within the range of a random distribution.

In summary, it turns out that the inclusive/exclusive opposition correlates significantly with singular and vertical homophony, but it does not correlate significantly with horizontal homophony. The two implications (U2) and (U3), that were extracted from the two significant correlations, point to the same direction. Singular and vertical homophony are preconditions for the presence of an inclusive/exclusive opposition. The person distinctions within the singular and within the non-singular have to be clearly separated before an inclusive/exclusive distinction is possible. A mix between singular and non-singular of

Table 4. Inclusive/exclusive opposition and horizontal homophony

	Inclusive vs. Exclusive		
	No	Yes	Total
No Horizontal Homophony	69 (26.0%)	65 (24.5%)	134
With Horizontal Homophony	75 (28.3%)	56 (21.1%)	131
Total	144	121	265

the same person (i.e. horizontal homophony) is indifferent to the presence of an inclusive/exclusive opposition.

These correlations can be understood as an effect of the essence of an inclusive and exclusive opposition. A first person plural is in fact a fusion of these two referentially highly different categories. Homophony between the inclusive and exclusive amounts to a form of indifference to the status of the addressee. In contrast, paradigms that mark an inclusive/exclusive opposition do not show this indifference; they mark 'pure person'. Pure person marking refers to the fact that paradigms with an inclusive/exclusive opposition consequently distinguish all different referential possibilities that can be made with the two most important interactional categories: the speaker and the addressee. The reference to 'speaker and others' is distinguished from the reference to 'addressee and others', but also from reference to a group that includes 'speaker and addressee'. The correlations as established in this section show that in a paradigm with an inclusive/exclusive distinction, the person reference is of such central importance that it is not possible to fuse person reference elsewhere in the paradigm. Only when the inclusive/exclusive opposition is not present, are other referential fusions possible.

5.   The Explicitness Hierarchy

The marking of pure person is better seen as a continuum between a more or less explicit marking of person. This continuum involves four different characteristics of paradigmatic structure. Three of these four characteristics have been discussed in the previous section: the inclusive/exclusive opposition, vertical homophony and singular homophony. The fourth characteristic that belongs to this continuum is the opposition between minimal inclusive and augmented inclusive. The minimal inclusive is a category that refers to the pairing of speaker and addressee.

The augmented inclusive is a category that refers to a group that includes the speaker, addressee and others. This opposition is a subdivision of the general inclusive category. These four characteristics form a hierarchy, as shown in Table 5. At the top of this Explicitness Hierarchy, paradigms maximally distinguish all eight referential categories by different morphemes. These paradigms have a minimal/augmented opposition, an inclusive/exclusive opposition, no vertical homophony and no singular homophony. At the bottom of the hierarchy are paradigms that barely distinguish person categories.

Table 5. The Explicitness Hierarchy

	Frequent Paradigmatic Structures					Infrequent Paradigmatic Structures				
Minimal vs. Augmented Inclusive	+	−	−	−	−	+	−	−	+	+
Inclusive vs. Exclusive	+	+	−	−	−	−	+	−	+	−
No Vertical Homophony	+	+	+	−	−	+	−	+	−	−
No Singular Homophony	+	+	+	+	−	+	+	−	+	+
Number of cases	26	78	99	20	21	3	12	4	1	1
	(244 cases, 92% of total)					(21 cases, 8% of total)				

This hierarchy can be thought of as showing more and more homophony from left to right, as exemplified in Figure 9. In the first stage, all eight person categories are distinguished. Starting from the first person complex, but invading the rest of the paradigms sequentially, more and more categories are taken together in the marking of one morpheme. In this figure, a prototypical example of each stage is shown. In the second stage of the hierarchy, the minimal and augmented versions of the inclusive are combined into one general inclusive category. Next, in the third stage, the inclusive and the exclusive are combined into a general first person plural. Only then is it possible for vertical homophony to occur. And just in case vertical homophony is present in the paradigm, singular homophony can eventually be found. Note that many possibilities for horizontal homophony exist, cross-cutting this hierarchy (not shown in Figure 9). Also, there are many more possible kinds of vertical and singular homophony than shown here in the last two stages of the hierarchy. The paradigms presented in Figure 9 are only prototypical exemplars of the various stages of the Explicitness Hierarchy.

As noted in section 2, in collecting the paradigms for this investigation, I have not *a priori* restricted the sample to either independent pronouns or
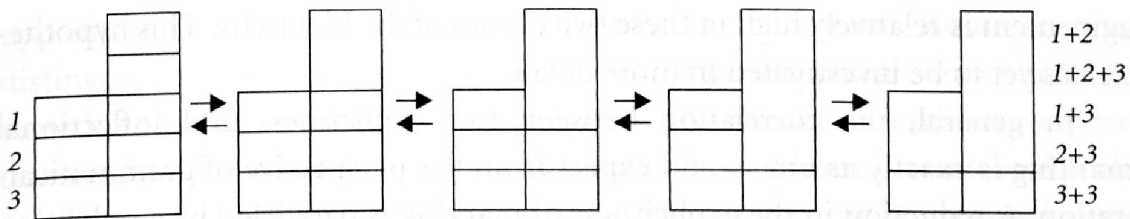
Figure 9. Exemplars of the Explicitness Hierarchy.

inflectional 'person agreement'. Including both kinds of paradigms in the current sample allows for an interesting conclusion. It turns out that the Explicitness Hierarchy is correlated with the morphological status of the paradigms. Highly explicit paradigms are more often independently marked; barely explicit paradigms are more often inflectionally marked, as shown in Table 6. The 'exceptional' cases, which fall outside the hierarchy, are more often inflectional than independent. Note that the mean percentage of inflectional paradigms over all 265 cases is not 50% but 55.8% (148 inflectional versus 117 independent).

Table 6. Morphological status of the Explicitness Hierarchy

	Explicitness Hierarchy					Others
Minimal vs. Augmented Inclusive	+	−	−	−	−	
Inclusive vs. Exclusive	+	+	−	−	−	
No Vertical Homophony	+	+	+	−	−	
No Singular Homophony	+	+	+	+	−	
Independently marked	21	41	42	10	0	3
Inflectionally marked	5	37	57	10	21	18
% Inflectional	19.2%	47.4%	57.6%	50.0%	100%	85.7%

The purer the marking of person, the more likely it is that the paradigm is marked as an independent morpheme. The differences are not very obvious in the middle range of the hierarchy, but to the extremes the skewing is quite clear. The paradigms high on the Explicitness Hierarchy are almost all independent pronouns, and the paradigms low on the hierarchy are all inflectionally marked. Surprisingly, in many cases low on the hierarchy, the paradigm is the basic way to refer to an argument — it is a 'pronominal argument' so to speak. The same holds for those inflectionally marked paradigms, which fall outside the hierarchy (labelled as 'Others' in Table 6). It is not necessary for these paradigms to be a form of agreement, although it might be the case that the proportion of

agreement is relatively high in these two classes of the hierarchy. This hypothesis has yet to be investigated in more detail.

In general, the correlation between low explicitness and inflectional marking is exactly as one would expect from the perspective of grammaticalisation. A reduction in the explicitness of marking is paralleled by a reduction in morphological independence. However, there is no indication that the hierarchy describes stages in a unidirectional change, as would be predicted by grammaticalisation. Changes in both directions of the hierarchy are possible (Cysouw 2001: 197–217). Yet, if a paradigm cliticises, then it is more likely that this paradigm will subsequently be reduced rather than be enlarged. This directional priority explains, at least partly, the skewing as found in Table 6. It can explain why there are more inflectional paradigms on the lower side of the hierarchy. However, it cannot explain why there are more independent paradigms at the higher side of the hierarchy. Furthermore, grammaticalisation cannot explain why horizontal homophony is not part of this hierarchy. Speculatively, I propose the varying level of awareness that speakers have of linguistic elements as a subsidiary explanation for these observations. Independent pronouns (like all independent words) are real things to a linguistically naïve speaker of a particular language. As speakers are consciously aware of the referential properties of independent words, unusual mixtures of referential categories are disfavoured. Highly explicit paradigms are, therefore, predominantly found as independent pronouns. At the other end of the spectrum, there is a lesser degree of awareness of the inflectional morphemes themselves, which opens up the possibility for unusual combinations of referential value for such elements, possibly as the result of some historical merger (cf. the history of the Germanic inflection). Consequently, paradigmatic structures that fall outside the Explicitness Hierarchy (labelled 'Others' in Table 6) show a large proportion of inflectionally marked cases. Horizontal homophony is not an unusual kind of mixture, as person marking remains constant; there is only a difference in number marking.

6.   Conclusion and prospects

The main finding of this cross-linguistic survey is the importance of the concept of 'pure person' marking. For pure person marking, all possible references to speaker and addressee (in different combinations) have to be distinguished in the person paradigm. The central characteristic of pure person marking is the

opposition between inclusive and exclusive first person plural. Paradigms that distinguish between inclusive and exclusive first person plural show much less variation than other paradigms. Paradigms with an inclusive/exclusive opposition rarely allow vertical homophony; singular homophony is not found at all.

The restrictions on the possible kinds of homophony can be captured in the Explicitness Hierarchy. This hierarchy is correlated with the morphological status of the paradigms. Paradigms that are highly explicit are more commonly found to be morphologically independent; paradigms that are barely explicit are more commonly found to be inflectional. Exceptions to this hierarchy also show a high percentage of inflectional cases. The correlation between the Explicitness Hierarchy and morphological status follows from grammaticalisation theory. Inflectional, i.e. highly grammaticalised, marking is less content oriented. Yet, historically the Explicitness Hierarchy does not seem to be followed unidirectionally — as would be expected from grammaticalisation theory. A subsidiary explanation for this difference in morphological status of the various paradigms can be found in the amount of conscious awareness people have of a linguistic element. Independent words are like 'things' to a language user. The conscious categories are structured by the difference in referential value, as this is important in the daily usage of a language. In contrast, inflectional categories are often opaque for the language user. In inflectional paradigms, a fusion of referential values is thus possible.

The Explicitness Hierarchy describes a continuum of more or less rich marking of person categories in a paradigm. Interpreted this way, the Explicitness Hierarchy could be a formalisation of the concept of 'richness', as it is used in the *pro*-drop hypothesis (cf. Harbert 1995:222). Only rich inflectional paradigms, so the hypothesis says, allow for *pro*-drop. Unfortunately, the notion of explicitness as developed in this paper does not seem to work as an explanation for *pro*-drop. Even on the lowest rung of the explicitness hierarchy, there are languages that allow for *pro*-drop (Cysouw 2001:51–53).

The generalisations as put forward in this paper could help to build a theoretical framework of linguistic generalisations that is informed by the diversity as attested in the world's languages. Both the usage of features (Ingram 1978; Noyer 1992:147–155) and the somewhat more powerful principle of feature geometries (Williams 1994:21–29; Harley and Ritter, this volume) are attempts to capture these regularities in the world's languages. Although nothing is wrong with such an attempt, the problem is that paradigmatic constellations that do not fit into such a framework will be described as accidental exceptions. However, it is still an open question as to whether there

is really a distinction between accidental homophonies and structural ones, "which cannot be ascribed to paradigm inconsistency, and which are not attributable to phonological factors either" (Carstairs-McCarthy 1998:330). From my investigations, it appears that the actual division between what counts as an accidental exception and what as a structural homophony is not clear-cut. From the generalisations put forward in this paper, a cline of predictability along the Explicitness Hierarchy can be abstracted. On top of the hierarchy is the minimal-augmented opposition.[7] If there is such an opposition between minimal and augmented inclusive, then the structure of a paradigm is strongly predictable — I would classify the few deviations as real accidental exceptions. If there is no minimal/augmented opposition but still an inclusive/exclusive opposition, then there are more deviations possible, and these deviations are partly internally regular. This means that these deviations are still exceptions, yet not as clearly accidental as in the case before. Further down, the possibilities of homophony between different referential categories become greater still. From this point onward, a homophony between first and second person, for example, is not an accidental exception anymore, but a structural possibility — although it would be an accidental exception when it would be found higher up the hierarchy. I leave it as a topic for future research to propose a framework in which such gradual regularities can be accounted for. For now, I want to make a plea for more attention to the (very large) actual diversity of the world's languages when developing a theoretical framework. A framework uninformed by the various linguistic possibilities may sound initially plausible, but is easily falsified.

Finally, I will present some other distinctions that seem to pattern with the generalisations presented. First, paradigms that distinguish number marking (dual, trial, paucal) seem to follow the Explicitness Hierarchy. Vertical homophony is rarely attested when there is an inclusive/exclusive opposition in paradigms with dual marking (Cysouw 2001:270–273). Second, gender marking is occasionally found in categories that include the speaker or the addressee. For instance, in Spanish, the first and second person plural have different forms according to gender (*nosostros* vs. *nosotras* and *vosotros* vs. *vosotras*). A more intricate example is found in the Semitic languages, which have gender in the second and third person singular and plural. In the Imperfect, the second person masculine is generally identical to the third person feminine (e.g. in Cairene Arabic, Gary and Gamal-Eldin 1982:100). More examples of gender marking in the first and second person can be found in Plank and Schellinger (1997). A major prediction that can be derived from the pure person hypothesis is that natural gender (in the first and second person) will not be found when

an inclusive/exclusive opposition is present in the paradigm. Gender marking does not qualify a paradigm for pure person. In a pure person paradigm only *person* marking counts. No other referential content (like gender characteristics) is allowed to mingle with this. There can only be gender marking in the first or second person if there is no inclusive/exclusive opposition. The only apparent counterexample to this claim is the Khoisan language Nama, but the gender marking and the inclusive/exclusive marking are probably two different systems (Haacke 1977). I do not know of any other counterexamples. However, this claim is open for falsification.

Appendix

The sample of the 265 paradigms that has been used for the analyses in this paper is a rather ad hoc collection of paradigms. The selection was primarily driven by the urge to collect as many as possible *different* paradigmatic structures. Because of this intention, some linguistic families are represented more often than others are. The 265 paradigms are collected from 234 different languages. From most languages, only one of the different available paradigms has been selected. A few languages turned out to have multiple paradigms of interest, all of which were included in this study. A few criteria have been used for the selection. First, every paradigm different from the nine common paradigms as shown in Figure 4 has been included. In contrast, I stopped collecting of these common paradigms after I had found 40 cases from all over the world. Also, paradigms with specialised number marking (dual, trial, etc.) and with (natural) gender marking in the first or second person are disregarded. Because of this decision, some linguistic families are underrepresented in this sample. Yet, the resulting collection is a good representation of the linguistic possibilities of person marking as attested among the world's languages. It might be the case that the amount of variation in this sample is slightly larger than would have been the case if I had used a perfectly balanced genetic sample of the world's languages. However, the claims that have been made in this paper would only have been much stronger if I had included all paradigms. The interesting variation would probably be lost in an even larger proportion of 'common' paradigms as shown in Figure 4.

The complete set of data is described elsewhere in detail (Cysouw 2001: Ch. 4). Due to reasons of space, I give here only a list of the linguistic families that are represented in this collection to assure that the sample consists of paradigms from a well-dispersed set of languages. The 265 paradigms come from the following families (ordered roughly by geographic region):

Africa: Niger-Congo (Adamawa, Atlantic, Bantoid, Dogon, Kru, Mande), Nilo-Saharan (Central Sudanic, Nilotic, Nubian, Saharan, Surmic), Afro-Asiatic (Semitic, Chadic, Cushitic, Omotic);

Eurasia: Indo European (Germanic, Romance, Celtic, Slavic, Iranian, Indo-Aryan), Uralic, Altaic (Turkic, Mongolian, Tungusic), Dravidian, Nakh-Dagestanian, South Caucasian, Chukotko-Kamchatkan;

Southeast Asia/Pacific: Chinese, Japanese, Tai, Ainu, Siau, Philippine Austronesian;
New Guinea: Sko, Isumrud, Central and South New Guinea, East New Guinea High-
                  lands, Binanderean, Suki, Central and South-eastern New Guinea,
                  Sentani, Border, Nimboran;
Australia: Bunaban, Nyulnyulan, Tiwi, Daly;
North America: Athabascan, Siouan, Salish, Muskogean, Sahaptin, Keres, Yuman, Algon-
                  quian, Miwok, Wakashan, Eskimo-Aleut;
Middle America: Uto-Aztecan, Oto-Manguean, Mixe-Zoque, Huavean;
South America: Chibchan, Choco, Gé, Arawakan, Carib, Aymaran, Quechuan, Tupí-
                  Guaraní, Tucanoan, Mascoian, Guiacuran, Paezan, Zamucoan,
                  Yanumam, Murã Pirahã.

Notes

* This paper has been written partly at the Institute for General Linguistics of the University of Nijmegen, and partly within the context of the project on pronominal clitics at the Zentrum für Allgemeine Sprachwissenschaft in Berlin. At those places respectively, I thank Leon Stassen and Paul Law for continued interest in these matters. Furthermore, I thank the present editors and two anonymous reviewers for their comments and additions to this paper. Special thanks go to Laura Downing for quick and thorough proofreading and for the many suggestions that substantially improved my English.

1. As a boundary case of person marking, I have included a few paradigms that have an opposition of the combination 'speaker/addressee' versus 'other', like the opposition zero versus -s in the English inflection.

2. Impersonal markers, reflexives, and reciprocals are excluded from this investigation.

3. Following Jelinek (1984), these are nowadays often referred to as 'pronominal argument' languages.

4. This kind of paradigmatic structure is also known as the 'Ilocano-type', referring to the influential analysis of the pronouns of the Philippine language Ilocano by Conklin (1962:135) — based on original observations by Thomas (1955). An earlier account of this phenomenon, which did not have any repercussions in the literature, is found in Foster and Foster (1948) for the Mixe-Zoque language Sierra Popoluca, spoken in Mexico.

5. Basically, it is argued there that multiple occurrences of 1 (i.e. speaker) and 2 (i.e. addressee) are not necessary to describe the attested cross-linguistic variation of person marking. This means that categories like 1+1, 2+2 or 1+2+2 do not have to be included in the frame as depicted in Figure 1. See also the discussion in *Studies in Language* by Greenberg (1988, 1989) and McGregor (1989).

6. The statistical measure 'Fisher's Exact' describes whether the distribution found differs significantly from a random distribution. For instance, in case of the distribution of Table 1, the cross-section of an inclusive/exclusive opposition and uncommon paradigmatic structures is found in 24 cases (9.1% of all 265 cases). The two parameters that are cross-sected at this

point in the table are attested respectively in 121 cases (45.7% of 265) and 81 cases (30.6% of 265). From these frequencies, it can be computed that — in a random distribution — the cross-section of these two parameters would be expected to occur in 37 cases (14.0% of 265 cases). Fisher's Exact now tells us that the actual number of 24 cases is only to be expected with a chance of .001 — which is small enough to warrant the claim that the deviation is not due to chance.

7. Note that the terms minimal and augmented are used here in the sense of McKay (1978), not in the sense of Harley and Ritter (this volume).

References

Borgman, D. M. 1990. "Sanuma". In *Handbook of Amazonian Languages*, D. C. Derbyshire and G. K. Pullum (eds), 15–248. Berlin and New York: Mouton de Gruyter.

Carstairs-McCarthy, A. 1998. "Paradigmatic structure: Inflectional paradigms and morphological classes". In *The Handbook of Morphology*, A. Spencer and A. M. Zwicky (eds), 322–334. Oxford: Blackwell.

Conklin, H. C. 1962. "Lexicographical treatment of folk taxonomies". In *Problems in Lexicography*, F. W. Householder and S. Saporta (eds), 119–142. Bloomington: Indiana University Press.

Cooke, J. R. 1968. *Pronominal Reference in Thai, Burmese, and Vietnamese* [University of California Publications in Linguistics 52]. Berkeley: University of California Press.

Cysouw, M. 2001. The paradigmatic structure of person marking. Ph.D. Diss., University of Nijmegen.

Forchheimer, P. 1953. *The Category of Person in Language*. Berlin: Walter de Gruyter.

Foster, M. L. and Foster, G. M. 1948. *Sierra Popoluca Speech* [Smithsonian Institution Publications 8]. Washington D. C.: U. S. Government Printing Office.

Gary, J. O. and Gamal-Eldin, S. 1982. *Cairene Egyptian Colloquial Arabic* [Lingua Descriptive Studies]. Amsterdam: North Holland Publishing Company.

Geary, E. 1977. *Kunimaipa Grammar* [Workpapers in Papua New Guinean Languages 23]. Ukarumpa: Summer Institute of Linguistics.

Greenberg, J. H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements". In *Universals of Language*, J. H. Greenberg (ed.), 73–113. Cambridge, MA: MIT Press.

Greenberg, J. H. 1988. "The first person inclusive dual as an ambiguous category". *Studies in Language* 12: 1–18.

Greenberg, J. H. 1989. "On a metalanguage for pronominal systems: A reply to McGregor". *Studies in Language* 13: 452–458.

Haacke, W. H. G. 1977. "The so-called 'personal pronoun' in Nama". In *Khoisan Linguistic Studies*, A. Traill (ed.), 43–62. Johannesburg: African Studies Institute.

Harbert, W. 1995. "Binding Theory, control, and pro". In *Government and Binding Theory and the Minimalist Program*, G. Webelhuth (ed.), 177–240. Oxford: Blackwell.

Harley, H. and Ritter, E. this volume. "Structuring the bundle: a universal morphosyntactic feature geometry".

Ingram, D. 1978. "Typology and universals of personal pronouns". In *Universals of Human Language*, J.H. Greenberg (ed.), 213–248. Stanford: Stanford University Press.

Jelinek, E. 1984. "Empty categories, case, and configurationality". *Natural Language and Linguistic Theory* 2: 39–76.

McGregor, W.B. 1989. "Greenberg on the first person inclusive dual: Evidence from some Australian languages". *Studies in Language* 13: 437–451.

McKay, G.R. 1978. "Pronominal person and number categories in Rembarrnga and Djeebbana". *Oceanic Linguistics* 17: 27–37.

Mühlhäusler, P. and Harré, R. 1990. *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: Blackwell.

Noyer, R.R. 1992. *Features, Positions and Affixes in Autonomous Morphological Structure* [MIT Working Papers in Linguistics]. Cambridge, MA: MIT Press.

Pence, A.R. 1968. "An analysis of Kunimaipa pronouns". *Kivung* 1: 109–115.

Plank, F. 1996. "Domains of the dual, in Maltese and in general". *Rivista di Linguistica* 8: 123–140.

Plank, F. and Schellinger, W. 1997. "The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45". *Linguistic Typology* 1: 53–101.

Thomas, D.D. 1955. "Three analyses of the Ilocano pronoun system". *Word* 11: 204–208.

Williams, E. 1994. "Remarks on lexical knowledge". *Lingua* 92: 7–34.