

The expression of person and number: a typologist's perspective

Author

Michael Cysouw
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
04103 Leipzig
Germany

Correspondence information

Email: cysouw@eva.mpg.de
Telephone: +49-160-94777076

Abstract

The categories of person and number have been analyzed extensively, both from a generative/structural perspective and from a typological/cross-linguistic perspective. The goal of both approaches is to account for the diversity of human languages, but in practice both have a rather different take on the subject. One major difference between these approaches is the relation between language-specific and cross-linguistic analyses. This paper argues that it is crucial to strictly distinguish between the two. However, this plea is at odds with the generative/structural perspective, which attempts to deal with both kinds of analyses at the same time. In contrast, it is of central importance from a typological/cross-linguistic perspective to keep the comparison constant across the wide variation as attested among human languages, thereby often ignoring many language-specific details (for the course of the comparison). The final section of this paper summarizes some major results of recent cross-linguistic comparisons of the person/number categories in the world's languages.

Keywords

person, number, typology, cross-linguistic, evidence, analysis

1. Introduction

When studying a grammatical phenomenon, like the expression of person and number, there is an important difference between language-specific explanations and cross-linguistic explanations (cf. Haspelmath 2007). I would even go as far as to claim that the search for these two kinds of explanations are completely independent endeavors, though they can (and should) inform each other. Ideally, the two undertakings will ultimately merge into one overarching theory of human language, but I do not think this juncture is near yet. As long as we are still searching for the right approach, it is important to clearly separate language-specific from cross-linguistic insights.

The difference between the two modes of inquiry is threefold. They have different scopes, different goals, and (at least for now) different explanatory modes. A language-specific explanation of the expression of person and number (or, for that matter, of any other grammatical characteristic) tries to find a unified account of as many as possible aspects of person-number marking in the single language under investigation. Every odd little aspect of the marking in this language is relevant and has to be considered. To model (or explain) all these aspects, normally an (elaborate) structural account summarizing the facts is deemed sufficient.

In contrast, a cross-linguistic perspective on the expression of person and number has a different scope. For a viable cross-linguistic investigation it is necessary to strictly (and restrictively) define the phenomena to be compared. This is important to assure that like is compared with like, thus not ending up comparing apples with dogs across different languages. In most empirical settings, such restrictions imply that not every detail of every language will be considered (and many otherwise interesting aspects of individual languages are deemed irrelevant). Also, the goal of a cross-linguistic investigation is different. Such a study will normally not try to cover each and every variant attested among the world's languages, but focus on the more widespread patterns occurring in many languages.¹ Regarding explanations, for a

¹ The assumption that cross-linguistic studies should only deal with widespread patterns is not necessarily shared by all researchers. There used to be a common assumption that cross-linguistic research deals with distinguishing possible from impossible languages (which in practice was often reduced to establishing a difference between attested and unattested languages). The empirical problems to differentiate possible from impossible is currently leading to more gradient approach to linguistic variation (see Section 4.2).

cross-linguistic theory it is central to explain why some variants are common, while others are rare (or non-existent). There will (probably always) be languages that do not fit into such a proposed cross-linguistic generalization, but such ‘counterexamples’ do not undermine the explanation, because cross-linguistic generalizations are necessarily probabilistic in nature (Cysouw 2003a, 2007a). Finally, the observation that a particular pattern is commonly attested among the world’s languages mostly evokes cognitive and/or functional explanations, though this is not a necessity.²

These differences between these two modes of linguistic inquiry, the language-specific and the cross-linguistics, are unfortunately rarely acknowledged. Ideally, these two approaches should of course merge into one theory of human language that would cover the insights from both kinds of investigations. However, in the search for such an overarching theory, I think it is important to remain aware of the empirical origin of insights to prevent unsuitable answers to good questions. As an example of a mix of arguments from both modes on inquiry, I will discuss the feature geometry for person/number marking number as proposed by Harley & Ritter (2002a) in Section 2. Next, in Section 3, I will present the basic decisions that have to be taken for a consistent cross-linguistic investigation of person/number marking, namely the choice of what kind of person/number markers are included in the comparison, and how is similarity between person/number markers established. Following on, in Section 4, I will present a survey of the recent flurry of cross-linguistic investigations into person/number marking. To completely review this vast literature would not fit into the current context, so I will focus on a few results that illustrate recent cross-linguistic insights into the marking of person and number.

2. The feature geometry by Harley & Ritter (2002)

The publications on person/number marking by Harley & Ritter (2002b, 2002a; dating back at least to an early manuscript by Ritter & Harley 1998) have had various follow-up discussion in the literature (e.g. Hanson 2000; Nevins 2003; McGinnis 2005; Benincà & Poletto 2005), but it seems that various problems with their ‘feature geometry’ approach are not widely acknowledged (though McGinnis 2005 points out some flaws, and proposes technical corrections). The study of person/number marking by Harley & Ritter (henceforth H&R) is truly cross-linguistic in

² The distinction between language-specific and cross-linguistic explanations is not in any way related to any variant of the formal-functional opposition. These two distinction are orthogonal to each other.

conception. For their research, they collected data on the world-wide distribution of independent pronouns (see the appendix to their 2002a paper). They claim that their model covers both general typological patterns and language-specific explanation, or, as they say, the geometry ‘gives us insight into both the distribution of pronominal paradigms in the world’s languages and the structure of unusual individual paradigms’ (Harley & Ritter 2002a:518). However, their account of the cross-linguistic variation is limited, and their notion of ‘unusual paradigms’ is strictly related to paradigms not fitting into their theory. Further, their feature-geometric model itself seems to have various shortcomings that will be discussed in Section 2.1. Independent of the explanatory model, it is also not clear what kind of data is deemed relevant. It appears that in principle every aspect of person/number marking from all human languages is relevant, though H&R are highly eclectic in the selection of data. This point will be discussed in Section 2.2.

2.1. Spelling out the feature geometry

There is an old tradition of analyzing person/number marking using features. The first explicit feature-like analyses of person/number marking were by Conklin (1962), Buchler & Freeze (1966) and Hollenbach (1970), just to mention a few early examples. Not much changed in subsequent installments of such feature-based analyses of person/number, including recent attempts like Dalrymple & Kaplan (2000) or Frampton (2002). In contrast, the model from H&R introduces various new ideas to the feature analysis of person/number marking. Most importantly, H&R introduce monovalent features and the bundling of features into geometries.

Even more innovative, H&R use the features not only to model specific categories as attested in individual languages (as has been done in all literature before), but also use them to make a typology of languages. Languages as a whole differ, they claim, as to which features are selected to be active in the language. This proposal clearly attempts to combine cross-linguistic generalizations with language-specific analysis. However, they only succeed by strongly restricting the cross-linguistic variation and by using only a rather limited selection of language-specific data to be accounted for. These are interesting proposals, but I will argue that they will not help us much for the modeling of cross-linguistic regularities of person/number distinctions in language, at least not in the way these proposals are used by H&R. Let me explain these points of criticism in more detail.

Monovalent features are features that are either present or absent, in contrast to the traditional binary features that are always present, but have either positive or negative value. This change might be useful for the interpretation of the features,

but it is important to realize that there is no difference in the modeling power between monovalent or bivalent features. Both kind of features offer two possibilities (present/absent or plus/minus), so a set of n features will lead to 2^n different feature combinations for both kinds of features.

The bundling of features into hierarchies is supposed to model the universal interaction between the features. However, every node in the hierarchy is itself a feature, which results in a proliferation of features. Consider a selection from the H&R feature hierarchy as shown in Figure 1. Instead of just four features (Speaker, Addressee, Group, Minimal) the hierarchical structure introduces two extra features (Participant and Individuation). Unordered, these six features would lead to $2^6 = 64$ different possible combinations. However, in the ordered 'geometrical' hierarchy, a feature can only be present when its parent in the hierarchy is also present, and this restricts the number of possible feature combinations.

Figure 1. Excerpt of the feature geometry from Harley & Ritter (2002a).

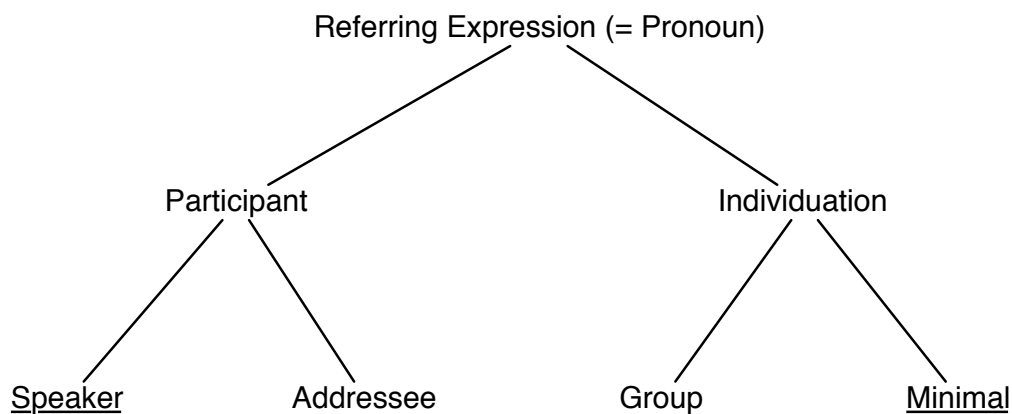


Table 1. Possible feature geometries.

Nr	Geometry	Example from H&R	Traditional analysis
1	RE	<i>hi</i> (Pirahã, p. 502)	3 number unspecified
2a	RE-indv	<i>me</i> (Daga, p. 489)	3 singular
2b	RE-indv-Min	<i>'aye'la</i> (Tonkawa, p. 492)	3 singular
3	RE-indv-Grp	<i>mu</i> (Daga, p. 489)	3 plural
4	RE-indv-(Min, Grp)	<i>'awe'la</i> (Tonkawa, p. 492)	3 dual
5	RE-part-Add	<i>gi(a)</i> (Pirahã, p. 502)	2 number unspecified
6a	RE-(part-Add, indv)	<i>ge</i> (Daga, p. 489)	2 singular
6b	RE-(part-Add, indv-Min)	<i>naya</i> (Tonkawa, p. 492)	2 singular
	RE-(part-Add, indv-Grp)	<i>e</i> (Daga, p. 489)	2 plural
8	RE-(part-Add, indv-(Min, Grp))	<i>wenaya</i> (Tonkawa, p. 492)	2 dual
9a	RE-part	<i>ti</i> (Pirahã, p. 502)	1 number unspecified
		<i>'üg</i> (Kwakiutl, p. 503)	1 singular
9b	RE-part-Spr	<i>-enuexu</i> (Kwakiutl, 503)	1 plural exclusive
10a	RE-(part, indv)	<i>ne</i> (Daga, p. 489)	1 singular
		<i>nee</i> (Winnebago, p. 512)	1 or 2 singular
10b	RE-(part, indv-Min)	<i>caya</i> (Tonkawa, p. 492)	1 singular
10c	RE-(part-Spr, indv)	<i>au</i> (Kalihna, p. 491)	1 singular
10d	RE-(part-Spr, indv-Min)	<i>yau</i> (Fijian, p. 495)	1 singular
11a	RE-(part, indv-Grp)	<i>nu</i> (Daga, p. 489)	1 plural (including dual)
		<i>geucaya</i> (Tonkawa, p. 492)	1 plural (without dual)
11b	RE-(part-Spr, indv-Grp)	<i>a?na</i> (Kalihna, p. 491)	1 plural exclusive
12a	RE-(part, indv-(Min, Grp))	<i>geucaya</i> (Tonkawa, p. 492)	1 dual
12b	RE-(part-Spr, indv-(Min, Grp))	<i>'eirau</i> (Fijian, p. 495)	1 dual exclusive
13	RE-part-(Spr, Add)	<i>-ens</i> (Kwakiutl, p. 503)	1 plural inclusive
14a	RE-(part-(Spr, Add), indv)	<i>kixko</i> (Kalihna, p. 491)	1 minimal inclusive
14b	RE-(part-(Spr, Add), indv-Min)	–	
15	RE-(part-(Spr, Add), indv-Grp)	<i>kixka:ro</i> (Kalihna, p. 491)	1 augmented inclusive
		<i>'eta</i> (Fijian, p. 495)	1 plural inclusive
16	RE-(part-(Spr, Addr), indv-(Min, Grp))	<i>'etaru</i> (Fijian, p. 495)	1 dual inclusive

Still, the selected part of the feature hierarchy in Figure 1 predicts the existence of 25 different person/number categories (including dual forms), and this is way too much for the world-wide diversity. For clearness of discussion, I have detailed out all 25 possibilities in Table 1, including some of the examples for the geometries that H&R present in their paper.³

As yet another technical notion to restrict the number of possible feature combinations (besides the tree structure), they also introduce the notion of default interpretation, which is only available for terminal nodes (Harley & Ritter 2002a:486). The default interpretations are indicated by underlining in Figure 1. Languages differ as to whether a node is activated by default, or explicitly specified. In Table 1, I have left out the default interpretation, but I grouped those geometries together that only differ according to spell-out or default interpretation. Note that the net-result of all these additional mechanisms is a restriction of the 64 theoretically possible categories to just 16 categories, which is of course also the number that one would get by taking four independent binary characteristics, like [\pm first], [\pm second], [\pm singular], [\pm plural].

So, the real reason to add all this technical mechanisms is not to make the feature system more expressive. The tree-geometric formalism is necessary to be able to model two aspects of linguistic variation at once, namely the different person/number categories themselves and the different systems of person/number categories as found in the languages of the world. In the view of H&R, languages differ as to whether they explicitly select features, or whether features are only spelled out by default (and they also add a third option, namely that the Individuation node can be absent). This leads to a typology of six different possible languages, as detailed in Table 2.⁴ The first four columns specify which features in the geometry are explicitly selected by the language. The fifth column gives an example language for each type, as discussed by H&R. As shown in the last two columns, this typology is far from original, and could easily be modeled with one binary feature [\pm clusivity] and one

³ H&R do not give any example for geometry number 23. Trying to follow their logic, this gap might be filled by a dual (or 'minimal') inclusive as found in a so called 'unit-augmented' systems, as first described for Rembarrnga by McKay (1978).

⁴ The number of six different language types seems to be far too few to come anywhere near the large diversity of person/number systems as attested among the world's languages (cf. Cysouw 2003b)

ternary feature like [no_number/only_plural/plural&dual]. Irrespective of the empirical adequacy of the approach by H&R (as discussed below), the main reason for the geometry is to account for 16 different possible pronoun categories and for the six different possible pronoun systems in one model. Now, that is a reasonable goal, but I do not think that it succeeds.

Table 2. H&R's Typology of languages based on explicit feature selection.

Speaker	Addressee	Group	Minimal	Example	Number	Clusivity
default	explicit	explicit	default	Daga	no dual	no
explicit	explicit	explicit	default	Kalihna	no dual	yes
default	explicit	explicit	explicit	Tonkawa	dual	no
explicit	explicit	explicit	explicit	Chinook	dual	yes
default	explicit	absent	absent	Pirahã	no number	no
explicit	explicit	absent	absent	Maxakali	no number	yes

There are various problems with the model of 25 feature geometries, as shown in Table 1, and with the typology of languages, as shown in Table 2. First, there are various geometries that have different possible meanings. The clearest example is geometry number 10a, which H&R explicitly discuss both for the pronoun *ne* from Daga, which is a first person singular pronoun, and for the pronoun *nee* from Winnebago, which is a pronoun with both possible first and second person reference. Comparable situations with multiple interpretations appear to exist for geometries 9a, 11a, and 15. H&R's approach to linguistic analysis assumes that one unified underlying representation can have different referential meanings on the surface depending on other factors active in the language. Although I personally do not agree with that approach, I acknowledge that this is a possible way to do linguistic analysis. However, I would expect that the mechanisms for the mapping of referential possibilities to a particular underlying geometry would be clearly laid out. To my knowledge, this is not the case in H&R's proposal (see Section 2.2 for more discussion of *ad-hoc* usage of data by H&R).

Second, there are many different geometries that seem to have fillings with the same meaning. For example, geometries 10a through 10d all are used for pronouns that are traditionally analyzed as first person singular. The differences between the

various first person singular pronouns as discussed by H&R is that they occur in different paradigmatic contrasts. However, there does not seem to be a reason to analyze the *first person singular* pronouns differently depending on whether, e.g., there is both a dual and a plural in the paradigm (as is the case for Tonkawa in geometry 10b) or only a plural (as is the case for Daga in geometry 10a). In contrast, it would make sense to have two different categories for the first person *plural*, depending on whether there is a first person dual or not. The referential value of a plural is special when a dual is available in the language, because plural then means ‘three or more’ and not ‘two or more’. However, this difference is not captured by the model of H&R. For example, the first person plural of Daga (without a dual) and Tonkawa (with a dual) are both analyzed using geometry 11a. It might be argued that the different first person singular geometries 10a through 10d are similar because they have identical geometries except for a difference between default interpretation or explicit spell out. However, in other cases this difference is judged to be intrinsically significant, for example in the case of 12a vs. 12b, the first of which is a first person dual, the second of which is a first person dual exclusive (cf. Cowper & Hall 2005:9-10).

Summarizing, various categories (like the first person singular) have different possible feature analyses in the model, although there is no reason why they should have more than one analysis. Conversely, some categories (like the first person plural) that might be given different analyses are not further distinguished by different geometries.

Turning now to the typology of languages of H&R (as summarized in Section 2.9 and the Appendix of their 2002a paper), there are likewise various inconsistencies. The system of Kalihna (Harley & Ritter 2002a:490) is a so-called ‘minimal-augmented’ system with eight different person/number categories including three different forms to be translated into English as ‘we’.⁵ This pronoun system is a clearly different from most other paradigm in their type ‘b’ (Harley & Ritter 2002a:520), which are ‘inclusive/exclusive’ systems that have only seven different person/number categories including only two different forms to be translated into English as ‘we’. Further, H&R list Kutenai in their type ‘h’, and the languages Perené Ashéninka Campa and Chalcatongo Mixteco in their type ‘i’ (Harley & Ritter 2002a:522), which actual-

⁵ This analysis actually depends on the interpretation of the suffix *-a:r-*, but that is a different story.

ly seem to be cases with absence of number marking.⁶ However, it is then unclear to me why the languages Maxakali and Kwakiutl are listed as belonging to type ‘b’, as H&R’s analysis (Harley & Ritter 2002a:502-503) seems to argue for them to be exactly of this type ‘i’. Summarizing, just as with the analysis of the pronoun categories, there are languages grouped together in one class that do not belong together and languages separated that should have been put into the same class.

So, although H&R’s goal is laudable, namely to combine a language typology and the analysis of individual categories into one structural account, I do not think that the specific proposals as made by H&R function properly. And then until now I have only considered the evidence that H&R themselves use. The interpretation of the feature geometry becomes even more difficult when further person/number markers from the world’s languages are considered (see the typological literature as discussed in Section 4 of this paper for more difficult cases). However, it is not completely clear what kind of evidence would count as relevant to their model.

2.2. What counts as evidence?

H&R explicitly invite falsification of their model: ‘the geometry proposed here makes predictions about possible and necessary contrast within a given system, as well as predictions about what constitutes a natural grouping of features. These predictions are falsifiable’ (Harley & Ritter 2002a:518). I would have preferred to discuss here the merits and shortcomings of their geometry on the basis of some particularly interesting examples.⁷ However, from the discussion in their paper, it is not particularly clear what kind of examples would count as (counter) evidence. The central question is whether syncretisms are informative or not, and, related to this

⁶ Again, this analysis depends on how one interprets the facts of these languages, but that discussion has to be reserved for another occasion. Further, note that Perené and Ashéninka are normally considered to be two different languages, and also note that there seems to be a typo in the appendix, as I think they meant to say for Kutenai that there is no number. It now says ‘Number sg/pl’, but I guess this should have read ‘Number: none (sg/pl)’, alike to the number designation of type ‘i’ directly below type ‘h’.

⁷ In reaction to a draft of the their 2002a paper, I already brought up some examples that did seem to be problematic, as can be seen from their discussion of Awa and Kawesqar (Harley & Ritter 2002a:513).

question, whether only independent pronouns are informative or also other evidence should be considered.

In the introduction to their paper, H&R declare that ‘paradigm-internal gaps and *syncretisms* are constrained by the hierarchical organization of features in the universal geometry’ (Harley & Ritter 2002a:482, *italic added*, MC). And again, in the conclusion, they state that the geometry is falsifiable by syncretisms because ‘a form in a paradigm could show syncretisms that demonstrate that some node would need to be dominated by the Participant and Individuation nodes simultaneously’ (Harley & Ritter 2002a:518). Both these citations indicate that H&R envision the occurrence of syncretisms to be informative for the establishment or falsification of their model. However, when checking for the use of the word ‘syncretism’ in the body of their paper, it turns out that they use it only for a different reason, namely to ‘explain away’ cases that do not fit into their geometry.

A clear example of this kind of argumentation is their discussion of the third person pronoun *ibisnó* from Koasati, which is underspecified for number, though there are number distinctions for the first and second person. For some unstated reason, H&R do not want to allow such a number split in their theory, so they offer to model the underspecification of number in the third person in terms of syncretism: ‘the different geometries are in fact present in the 3rd person, but the morphological inventory of the language does not provide distinct forms to realize them’ (Harley & Ritter 2002a:510). Here the term syncretism is used to describe a situation in which two different geometries should both exist in the language, but they have the same surface form and are thus superficially indistinguishable. H&R argue for the existence of both geometries in Koasati by offering other evidence from the same language in which the difference is made.

This is a clear case of ad-hoc mixing of language-specific argumentation into a cross-linguistic investigation, which I think is not viable. I find that these kind of structures as found in the world’s languages (e.g. a number split as attested in Koasati) should inform the theory instead of be treated as a nuisance. It is the world’s attested diversity that helps us refine and redefine our theories. Now, there is a practical question of course where to stop. One could go on forever looking for yet other structures in yet other languages before finally building a theory. In a sense, H&R did their best in compiling a large collection of pronoun paradigms on which to base their theory. They then tried to account for new and unexpected cases, also if their theory was not made to handle them. Unfortunately, there is a lot of published literature about ‘strange’ pronoun systems, most of which could easily have been included by H&R (Cysouw 2001b is cited by Harley & Ritter 2002a,

though they do not discuss the many unusual paradigms present there. The knowledge about the world's linguistic diversity has only been growing since then, cf. Section 4).

The same argumentation as used for Koasati is used for all examples that do not readily fit in their model, viz. the examples from Winnebago (Harley & Ritter 2002a:512), Awa (Harley & Ritter 2002a:513) and Biak (Harley & Ritter 2002a:517-518). Apparently, syncretisms are not necessarily informative for the establishment of the theory itself. In contrast, when discussing the link between the inclusive and the second person, H&R are already satisfied that in a few cases 'the inclusive form patterns morphologically *in the prefix* with the second person form' (Harley & Ritter 2002a:504, italics added, MC). The whole forms of the inclusive and the second person pronouns are clearly different in all the cases discussed by H&R. Apparently, partial syncretisms already counts as evidence when they fit into their model (see Cysouw 2005c for an extensive discussion of inclusive-second person syncretisms in the world's languages).

Further, H&R do not strictly follow a particular strategy to delimit which markers should be included in their study. Originally, H&R collected paradigms of independent personal pronouns from a sample of the world's languages, and all examples they give in their paper are independent pronouns. Further, they use the word 'paradigm' almost on every page of their paper, so my first impression was that they set out to model the person/number categories as present in the world's paradigms of independent pronouns. One might criticize a restriction to only include independent pronouns in the study, but such a restriction certainly gives a clearly defined subdomain to investigate from a cross-linguistic perspective (cf. Cysouw 2003b: 32). However, after discussing Pirahã, H&R suddenly introduce another concept: 'we argue that these rare languages [like Pirahã, MC] are in fact without 'true' number distinctions. [...] These languages are to be distinguished from languages that do not express number distinctions in pronouns but have such distinctions elsewhere, typically on common nouns or in verbal agreement' (Harley & Ritter 2002a:502). So, markers elsewhere in the language also count as evidence. However, other evidence is only taken into account when there are problems to fit a particular set of independent pronouns into the feature geometry (cf. Harley & Ritter 2002a:510-513). Apparently, independent pronouns are sufficient evidence as long as they fit into the model. Only if they do not fit, then other evidence from the language is considered relevant.

This kind of argumentation clearly brings up the basic disagreement between most typological work on the world's linguistic diversity (including mine) and many

structural approaches to linguistic analysis (including H&R's approach). The basic question is what kind of evidence should be included in an investigation: is just every bit of information from every language always noteworthy, as H&R seem to assume, or should one delimit the kind of evidence for the course of each study, as I would argue. From my experience with the world's linguistic diversity, one should worry that by opening up the doors for all evidence available there is a obvious and very deep trap of (unintended) cherry-picking. By stumbling from one interesting fact to the next it is difficult to keep track of all the other possible observations that one did not consider, but could have. To lay out the linguistic diversity to be explained by any theory of human language, I think it is crucially important to keep the comparison constant. To keep the comparison constant, it is important to strictly adhere to a particular decision to include only one strictly defined kind of markers (e.g. independent pronouns) and not use other markers opportunistically (as discussed in the next Section 3.1). Further, syncretisms (or 'language-internal similarity between markers') are important observations to build cross-linguistic generalizations. However, also here it is important to adhere to one's definition of what counts as a case of syncretism, and not make ad-hoc judgements on the basis of idiosyncratic facts from individual languages (cf. Section 3.2).

3. Keeping the comparison constant

For a consistent cross-linguistic comparison, it is necessary to decide: (a) which data is relevant for the comparison; (b) how similarities are established *across* languages; and (c) how similarities are established *within* languages.

The first question (a) will be detailed in Section 3.1. The rather obvious point is that there are different kinds of person/number marking in human language (e.g. morphologically independent markers vs. inflectionally bound markers, or subject cross-reference markers vs. possession markers, etc.). The choice of markers influences the results of any cross-linguistic comparison.

The second point (b) is a central problem for all language comparison: how to equate any construction from language A with an construction of language B? To make consistent comparisons, some kind of language-external *tertium comparationis* is necessary. Fortunately, for person/number marking this is not so much of a problem, because their deictic nature. For example, it is relatively straightforward to equate the notion 'speaker' across languages (though for a different view see Goffman 1979).

The third issue (c) is more interesting, and I will discuss this topic extensively in Section 3.2. The central input to any cross-linguistic theory about person/number

marking are languages in which particular meanings are combined into the reference of one particular form (cf. the discussion of syncretisms in Section 2.2). Formulated more comprehensively, generalizations about the structure of person/number marking in human language depends on the establishment of similarities between person/number markers within a language. Syncretism is simply an extreme kind of similarity, namely complete identity. The more often two meanings are coded similarly in the world's languages, the more similar the meanings. Or, conversely, when two meanings are never coded identically, then they should be considered completely different by any theory (see Cysouw 2010 for the central importance of language-internal similarity for cross-linguistic research). However, there are different possibilities how to establish such similarity between markers, and this has strong repercussions on the resulting theory.

3.1. Person/number marking is more than personal pronouns

In the investigation of the category of person/number in language there is a tradition to restrict the research to personal pronouns. Examples of this limitation are Chlenova (1973), Laycock (1977), Sokolovskaya (1980), Schwarz (1986), Harley & Ritter (2002a), Daniel (2005) and Bhat (2004), only to mention some of the typologically oriented investigations into person/number marking from recent decades. Although such a restriction does not include all possible person/number markers as found in human language (and thus might be considered too restrictive), it is important to realize that it is at least a consistent sampling technique. In contrast, Forchheimer (1953) was—rightly I think—criticized for not consistently sampling the same kind of markers (for a discussion of Forchheimer's contemporaneous criticism, see Cysouw 2003b:25-28). Also, I have criticized Harley & Ritter (2002a) above in Section 2.2 for not holding on strictly to their original data, which only consisted of independent pronouns. By adding examples of inflectional person/number markers on an ad-hoc basis they abandoned a clear data-selection strategy. It is of course not necessary to only investigate independent pronouns (see for example Baerman et al. 2005 for a more all inclusive approach), but whatever decision is taken on what kind of data are deemed relevant, that decision has to be kept constant until any conclusions are taken.

There are some indications that it actually matters what kind of person/number markers are considered. First, there are typological differences depending on whether the person/number markers are morphological independent or inflectional-

ly bound.⁸ However, only when both person/number inflection and personal pronouns are included in the investigation is it of course possible to decide whether there are any differences between them. Cysouw (2003b:311-315) finds various structural correlations, all indicating that independent pronouns tend to have more different person/number markers than inflectional person/number markers (which is completely in line with expectation from the perspective of grammaticalization). Siewierska & Bakker (2005:158-160) investigated marking differences between inflectional and independent marking for the various fine-grained clusivity distinctions. For this domain, they conclude not to find any clear correlations between the person-marking structure of subject pronouns and subject inflection across languages (in contrast to the claims by Cysouw 2003b). However, doing just a very basic comparison between the data from the two papers the same preference seems to be discernible (cf. Table 3, column “% Independent”, computed by taking the fraction of independent pronouns relative to all cases).⁹ Summarizing, independent person/number markers, in contrast to inflectional person/number markers, have a preference for more semantic differentiation (see also Siewierska & Bakker 2006).

⁸ With ‘inflectionally bound’ I only refer here to main argument cross-referencing that is marked morphologically bound to a main predicate. This is often referred to as ‘agreement’, but I find that term unfortunate, because it assumes that independent pronouns are somehow primary to inflectionally marked person markers, which is cross-linguistically not the most widespread situation. The analysis of clitics—being intermediate structures between independent pronouns and inflectional person/number markers—is of course important in this context, but will not further be discussed here.

⁹ To compare the data from Cysouw and Siewierska and Bakker, I combined the labels ‘vertical homophony’ and ‘singular homophony’ from Cysouw (2003b:311) into the label ‘no we’, and the labels ‘only inclusive’ and ‘inclusive/exclusive’ from Siewierska & Bakker (2005) into ‘inclusive/exclusive’. The minor categories are ignored here.

Table 3. Comparison of Cysouw (2003b) and Siewierska & Bakker (2005), showing both a preference for less person/number oppositions in inflectional person/number marking.

	Cysouw (2003)			Siewierska & Bakker (2005)		
	Independent	Inflection	% Indep.	Independent	Inflection	% Indep.
no we	17 (12%)	43 (30%)	28.3%	8 (2%)	26 (9%)	23.5%
unified we	42 (30%)	63 (43%)	40.0%	224 (59%)	166 (57%)	57.4%
inclusive/exclusive	47 (34%)	32 (22%)	59.5%	132 (35%)	91 (31%)	59.7%
minimal/augmented	32 (23%)	7 (5%)	82.1%	15 (4%)	8 (3%)	65.2%
Total	138	145		379	291	

Looking closer at inflectional person/number marking, Cysouw (2001a, 2003b:315-318, 2009) finds a striking asymmetry between person/number prefixes and person/number suffixes. In general, paradigms of person/number prefixes are much smaller than paradigms of person/number suffixes. More concretely, prefixes much more often show syncretisms between singular and plural ('horizontal homophony' in the terminology of Cysouw 2003b). In some—but far from all—of these cases there are (optional) disambiguating suffixes, but in far from all. So, it does not seem to be sufficient to simply distinguish independent pronouns from inflectional person/number marking. Even the kind of inflectional marking makes a difference as to the kind of syncretisms attested.

Some preliminary observations suggest that functionally different person/number markers also show special marking effects. For example, zero marking is extremely uncommon cross-linguistically for the first person singular in inflectional main argument person/number paradigms (the more so with co-occurring overt second and third singular). However, in the person marking of kin possession ('my father') zero first person marking seems to be much more widespread (Cysouw 2003b:58-59). Another case of special marking of person categories is found in imperative/hortatives, in which addressees are the unmarked persons, leading to special distributions of zeros in the paradigm, and to a preference of first person plurals to have an inclusive meaning (Cysouw 2003b:60; Dobrushina & Goussev 2005). So, in summary, the choice of person/number markers will clearly influence the result of any cross-linguistic investigation.

3.2. Similarity between person/number markers

When comparing person/number markers—or any other kind of markers for that matter—there are various levels of linguistic structure that might be deemed relevant. Choosing one or another of these levels is not a question of right or wrong—all have their merit—but they should not be confused. Any comparative decision of what counts as similar should be followed throughout.

There is general sense in which language-internal similarity is at the heart of linguistic theorizing. When two apparently different meanings/functions are expressed similarly—or even identically—in a particular language this raises the question why this should be the case. Now, there are many different ways to explain such a situation, which I will not further discuss here. Yet, the important prerequisite for such an observation of language-internal similarity is a definition of what counts as similar. The notion ‘syncretism’ is often used to express such a notion of resemblance, though the underlying definition of similarity is mostly not made explicit. This section explores different possible definitions of language-internal similarity as used for the establishment of a theory for person/number markers. Such similarity within a language can at least be established on four different levels: (i) submorphemic, (ii) morpheme-based, (iii) wordform-based, and (iv) language-based.

In submorphemic analysis, two person/number markers are considered to be similar when they share some phonemic similarity, possibly of a smaller extent than a traditional morpheme.¹⁰ For example, there is a clear sense in which German *mich*, *mir* and *mein* share something (they are all first person singular pronouns, and they all start with \m\), though traditionally there is no German morpheme *m-* ‘first person singular’. Such proposals to submorphemic analysis go at least back to Bolinger (1950). An early attempt to use this approach for the analysis of Germanic pronouns is Pike (1965). For a summary for many similar attempts for Germanic, see Howe (1996:32-42). The most common situation in which such an analysis is applied is to separate person from number (cf. Forchheimer 1953; Daniel 2005). More recently,

¹⁰ To be submorphemically similar, the markers also have to be similar in functional/semantic terms. Actually, it is the combination of both functional formal similarity that leads to a submorphemic analysis, cf. the ‘syncretism principle’ from Müller (2006a:163). Also note that the difference between submorphemic and morphemic is often difficult to establish (and it might even be impossible in general). For the discussion here I use a rather traditional notion of what counts as a morpheme, and everything smaller than that is called ‘submorphemic’.

submorphemic approaches to the similarity of person markers have been revived and greatly extended by G. Müller and his co-workers in the context of distributed morphology (Müller 2006b; Fischer 2006; Georgi 2006).

More traditionally, two person/number markers are only considered to be similar when they share complete morphemes. This is actually the default approach in most typological studies into person/number marking. From this perspective, the German pronouns *mich*, *mir* and *mein* are simply different, and their partial similarity is ignored. Although this might seem to be missing an important generalization for German, the clear advantage of the morphemic approach is that it is easier to keep the level of comparison constant across languages. For example, the English pronouns *me*, *he*, *she*, and *we* also share a phoneme, and might be analyzed as showing some submorphemic similarity. Of course, this collection of person/number categories is semantically much less coherent than the German case (*me*, *he*, *she*, and *we* combine first person singular and plural with third person singular, and mixing subject and object forms), but it is rather difficult to consistently draw the line between a coherent and a haphazard semantic similarity across a wide range of languages.

Concerning inflectional person/number markers, there are two slightly different interpretations of the morphemic approach to the similarity between markers. Cysouw (2001b, 2003b) separates non-contiguous inflectional morphemes into different paradigms. For example, person/number inflection in Svan (a South Caucasian language) consists of both prefixes and suffixes (Tuite 1997:28). The set of prefixes and the set of suffixes are considered as separate paradigms by Cysouw (2003b:11-12). This approach results in some similarity between, for example, first person singular *xw-...-äs* and first person plural exclusive *xw-...-ad*, because both have the same prefix (but not the same suffix). Different from the submorphemic approach, Cysouw only separates person/number markers into parts when they are clearly attested as separate 'slots' in the morphemic structure of the language in question.

A different take on morphological similarity is taken by Baerman et al. (2005). In their perspective, two person/number categories are only similar if the complete wordform is identical. So, taking the above mentioned inflectional forms from Svan as an example, Baerman et al. will consider them to be simply different. Again there is the question as to whether this wordform perspective misses some generalizations. However, it is a clear and consistent strategy to investigate similarities between person/number markers. Further, the wordform approach circumvents the problem of consistently identifying separate 'slots' of person/number morphemes across languages with complex morphological structure.

These first three levels of measuring similarity operate within a paradigm of person/number markers. However, a single language normally has various paradigms of person/number markers (as discussed in the previous section). The most restrictive measurement of similarity between person/number markers is when only those similarities count that are found in all paradigms in the whole language. McGinnis (2005:706ff) calls this ‘conflation’ (in opposition to ‘syncretism’ that is restricted to similarities between markers in incidental paradigms in a language). Conflation between person/number markers is much less common, and it will thus be much more difficult to find enough evidence for any cross-linguistic conclusion. Moreover, there seem to be various rather unexpected cases of conflation—with ‘unexpected’ in this context meaning that examples of conflation are attested for extremely rare syncretisms.

For example, in Yagua (an isolate from Peru) all person/number paradigms distinguish between dual and plural forms. However, all these paradigms also have just one inclusive form, which does not distinguish dual from plural (Payne & Payne 1990:361-371). Inclusive with second person plural conflation is attested in Sanuma (a Yanomam language from Venezuela, Borgman 1990:149), though such a syncretism is extremely uncommon from a morphemic perspective. Next, Gooniyandi and Bunuba (two closely related Bunaban languages from Australia) have one form for the augmented inclusive (‘you and me and some other’) in opposition to a form that combines minimal inclusive (‘you and me’) and exclusive (‘I and other’). This is an extremely uncommon combination, but it still permeates throughout the structure of all person/number paradigms in both these closely related languages (McGregor 1990:167-169, 1996; Rumsey 1996, 2000:70-72, 80-82). Exactly the same structure is also attested in Yaouré (a Mande language from Côte d'Ivoire), likewise found throughout all paradigms of person/number marking in the language (Hopkins 1986). A different take on these ‘we-like’ categories is found in Guató (a Macro-Ge language from Brazil, Palácio 1986) and Pech (a Chibchan language from Honduras, Holt 1999). Both these languages have one form for minimal inclusive (‘you and me’) in opposition to a form that combines reference to exclusive (‘I and other’) with augmented inclusive (‘you and me and others’). Again, these unusual structures are found throughout all paradigms of person/number in these languages. For the complete argumentation that these structures are rare, see Cysouw (2005c).

This difference between structures as attested with conflation and those attested with syncretism shows that it is important for a cross-linguistic study to strictly adhere to a particular measure to establish language-internal similarity between per-

son/number markers. Depending on the measure chosen, different theoretical generalizations might be needed to explain the cross-linguistic preferences.

4. Cross-linguistic insights on person/number expression

In recent years, a great number of monograph-sized studies of the structure of person/number marking from the perspective of the world-wide linguistic variation has appeared (Baerman et al. 2005; Bhat 2004; Cysouw 2003b; Filimonova 2005b; Gehling 2004; Helmbrecht 2004b; Siewierska 2004). An even larger number of highly data-driven studies have been published as articles and data collection.¹¹ I will not give an impartial summary of this whole literature, but focus on a few aspects that are closest to my own interest. First, I will discuss the special relation between person and number, when looking at it from a cross-linguistic point of view. I will argue that from this perspective it seems to be best to consider person and number as one parameter, and not as the intersection of two independent features ‘person’ and ‘number’. This is not necessarily the best analysis for each languages, but it makes most sense for the overwhelming majority of them (Section 4.1). Further, I will look at attested restrictions in the cross-linguistic variability of person/number markers. Though not everything imaginable is attested in the world’s languages, I will argue that the difference between attested and unattested is less important than the difference between frequent and rare. Such a view implies that any theory about the cross-linguistic variability has to be probabilistic in conception (Section 4.2).

4.1. The special relation between person and number

There is an old tradition to analyze personal pronouns by a cross-section of person and number features. This approach goes back all the way to the origin of occidental linguistics (starting with Dionysios Thrax, cf. Uhlig 1883; Swiggers & Wouters

¹¹ For everybody interested in the full depth of the recent typological studies into the structure of person/number marking, I consider the following papers and data collections to be essential reading, restricting myself to publications since 2002 (for earlier works, please consult the references as discussed in these papers): Baerman 2002, 2004, 2005; Baerman & Brown 2005; Baerman et al. 2005; Bhat 2004, 2005; Cysouw 2003b, 2005a,b,c,d, 2007a; Daniel 2005; Daniel & Moravcsik 2005; Dobrushina & Goussev 2005; Filimonova 2005a; Gehling 2004, 2006; Helmbrecht 2004a,b; Siewierska 2004, 2005a,b; Siewierska & Bakker 2005; Simon 2005.

1998). However, at least since Boas (1911:39) and Jespersen (1924:192) it has been observed that the notion ‘plural’ in the person-marking domain seems to be something different from the plural as found with nouns. For example, the English pronoun *we* normally does not mean ‘plural I’, which would literally be reference to a group of people speaking in unison (a so-called ‘choral we’ as sometimes found in situations of mass speaking). The pronoun *we* can have such a meaning, but this usage is only extremely rarely attested. Normally, English *we* refers to a group of people consisting of the speaker together with some non-speakers. Such special semantics of plurality is also attested with nouns, a phenomenon that is in the literature referred to as an ASSOCIATIVE plural (Moravcsik 1994, 2004). The same semantic analysis also holds for the second person plural, which would literally only refer to ‘present audience’, but in practice mostly refers to one addressee with various others (for an in-depth discussion cf. Cysouw 2003b:74-77).¹²

Not only are the ‘plurals’ of the person marking semantically different from regular nominal plurals, the difference between the plurals is also attested in morphological structure. First, as far as we know, there is no clear case of a grammaticalized ‘real’ plural of the first or second person in any person marking system in the world’s languages (i.e. there is never a specialized form for the ‘choral we’ or the ‘present audience’ meanings, cf. Cysouw 2003b:73-74; Simon 2005). Second, it is highly unusual to find transparent morphological ‘plural’ inflection in the first and second person. In a sample of 260 languages, Daniel (2005) finds 42 languages in which there is an apparently regular morpheme deriving the first person plural from the first person singular (16%, though including eight Altaic and six Sino-Tibetan languages, so this 16 % is probably an overestimate). In only 19 of these 42 languages this morpheme is the regular nominal morpheme (7%, including all six [!] Sino-Tibetan languages).¹³ From a discussion of a few of the other 23 examples in

¹² Recent proposals for the semantic analysis of plurality attempt to unify these different kinds of plurals (cf. Sauerland 2003). Although this unification is technically perfectly possible, the main empirical question remains whether there is still a need to differentiate between different kinds of pluralities. In most cases, the meaning of the plural phrase can be predicted from the head noun, though not in all cases. For example, Moravcsik (2004:493) mentions the case of Japanese *kodomo-tachi*, which can mean both ‘the child and his associates’ and ‘more than one child’.

¹³ Daniel (2005) reports on another 22 languages in which a nominal plural marker is found with first and second person plural pronouns, though in these cases these

Cysouw (2003b:116-117) it is clear that these apparent 'plural' morphemes are often related to associative markers in the languages in question.

These observations that plural pronouns are rather unlike nominal plurals led Cysouw (2001b, 2003b) to propose a redefinition of the notion of 'plurality' in the domain of person marking. The claim is that the traditional notion of 'plural' has no place in a cross-linguistic theory of person/number marking (though it might conceivably be of use in the third person). In contrast, higher numbers (like dual, trial, etc.) are real number dimensions (both judging by their semantics and by their formal construal in the languages that have them, cf. Cysouw 2003b, Ch. 6). Looking closer to the possibilities within the domain of singular and 'plural' person marking, Cysouw proposes the existence of eight different person/number categories, as summarized in Table 4. Three of these categories involve only one participant, so they can be called 'singular'. The remaining five categories involve more than one participant, so they can be called 'non-singular'. This distinction was used in the tabulation of examples in Cysouw (2001b, 2003b). However, this distinction is rather superficial to the analysis of person/number markers (as I have just argued). This superficiality is also clearly indicated by person/number paradigms that distinguish all eight primitives, for which it has been argued ever since Thomas (1955, i.e. more than fifty [sic!] years ago) that the minimal inclusive ('dual inclusive') behaves somewhat alike to a singular, though it is of course is not a singular category in a semantic sense.

markers only function as reinforcements as the stem forms of the singular and the plural person markers are already differentiated.

Table 4. person/number primitives following Cysouw (2003).

Name	Number of participant	Old abbreviation	Extendable	New abbreviation
speaker	one	1	no	1
addressee	one	2	no	2
other	one	3	no	3
minimal inclusive	more than one	1 + 2	no	12
augmented inclusive	more than one	1 + 2 + 3	yes	12 +
exclusive	more than one	1 + 3	yes	1 +
addressee with others	more than one	2 + 3	yes	2 +
others	more than one	3 + 3	yes	3 +

In retrospect, it was probably not clearly enough stated in Cysouw (2001b, 2003b) that the abbreviations used for the eight categories (cf. the third column in Table 4) are not supposed to be a feature-like analysis of person/number marking. The abbreviations should be seen more like a mnemonic device to quickly see which category is referred to in a discussion, than it was supposed to be a proposal for a further breakdown of the primitives. It might also have been possible to use eight arbitrary names, like *a* through *h*, or p_1 through p_8 . Such abbreviations clearly miss out on some of the obvious semantic structure available. However, what *exactly* are the semantic possibilities of each of these categories has not been conclusively discussed in Cysouw (2001b, 2003b).

If a further disassembly of the eight primitives is wanted, it seems of most interest to me to introduce a notion of ‘extendibility’ (cf. the fourth column in Table 4). With this I mean that there are some person/number categories that can be defined by a finite listing of participants. The non-extensible categories are *speaker*, *addressee*, *other* and *minimal inclusive*. The other four categories are the extended versions of these four categories, meaning that some participants have to be added, either one of the kind of participants already present, or any ‘other’ or ‘third person’ participant. So, for example, the exclusive (‘1 + 3’) category could consist of combination like 1 + 3, 1 + 1 + 3, 1 + 3 + 3, etc. To more clearly indicate this notion in

the abbreviations used, I propose some new abbreviations in the last column of Table 4, using a '+' to indicate possible extension of the reference.¹⁴

Most importantly, this proposal is not just a graphical revision, but it actually represents a hypothesis for further cross-linguistic research, because I am not sure that the internal semantics implied holds for all (or at least the large majority) of the world's languages. The first consequence of this hypothesis is that the 'choric *we*' usage (1 + 1, i.e. various speakers speaker at the same time) should always be coded by the exclusive form, 1 +. This is the only form that would allow for the meaning 1 + 1, following the above outlined hypothesis. It would be particularly interesting to test this in languages that have an inclusive/exclusive distinction. A counterexample to this hypothesis would be a language in which the choric *we* is expressed by an inclusive form. Likewise, the 'present audience' usage (2 + 2, i.e. addressing various people who are all present in the audience at the time of the speech act) should always be coded by the same form that also codes for addressee with others, 2 +.¹⁵ Finally, different combinations of various speakers and addressees, but without others (e.g. 1 + 1 + 2, 1 + 2 + 2, 1 + 1 + 2 + 2, etc.) should all be coded alike to the augmented inclusive, 12 + (cf. Bobaljik 2008:219-220). As a rationale behind this hypothesis I would like to propose that in the conceptualization of human interaction there is just one speaker and one addressee. Other speakers in unison or multiple addressees are always metonymically reduced to a single entity (cf. Cysouw 2003b:77). Multiple unison speakers 'speak as one' and multiple simultaneously addressed people are 'addressed as one'.¹⁶

¹⁴ Specific number categories in the person domain, like dual, trial, or paucal (called 'restricted' numbers in Cysouw 2003b) can be denoted by adding subscripts to the plus-sign, indicating how many possible 'plusses' there are. For example, an second person trial would then be 2 +².

¹⁵ Simon (2005) argues that 2 + 2 is never grammaticalized in human language (at least, there is no good known example for it). However, he does not delve into the question how languages codify this category, which is semantically perfectly possible.

¹⁶ For testing all these hypotheses, it is important to keep in mind the many situations in which social forces will override the referential possibilities, as is particularly prominent in social settings that induce honorific language use (cf. Cysouw 2005a on honorific usage of inclusive and exclusive forms). Further, the idea that choric *we* is closely connected to exclusive pronouns might be true referentially, but that is

4.2. On the limits of variability

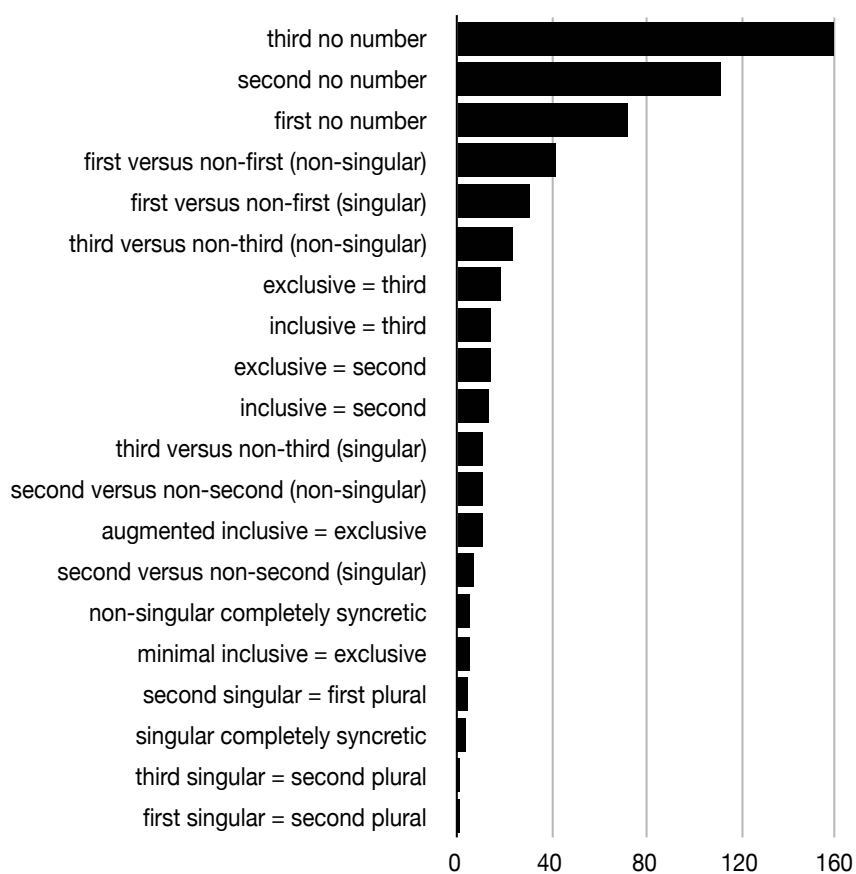
Finally now, all parts of the cross-linguistic puzzle can be put together. On the basis of (i) a set of cross-linguistic categories (see Section 4.1), (ii) a decision on the cross-linguistic selection of data (see Section 3.1), and (iii) a decision on how to establish similarity within each language (see Section 3.2), it is possible to make a survey of syncretisms between cross-linguistic categories. Any theory or model of language can then be tested as to its applicability with regard to these syncretisms attested. The better a theory can predict the attested frequencies of syncretism, the more suitable the theory.

For example, given the eight person/number categories as summarized in Table 4, and given the cross-linguistic data collected by Cysouw (2003b), the frequencies of person/number syncretisms can be compiled. The most commonly attested combinations are ‘inclusive’ (i.e. 12 and 12+ combined) and first person plural (12, 12+ and 13+ combined). The frequencies of all remaining syncretisms as described in Cysouw (2003b) are summarized in Figure 2.¹⁷ An important observation that arises from the frequency distribution in Figure 2 is that the distinction between attested and unattested seems to be rather vague. There are many unusual syncretisms that are attested in just a few cases. The difference between a syncretism being attested once and an unattested syncretism will very strongly depend on the specific details of the sample of languages investigated, on the world's languages that happened to be described in any detail, and on the languages that happened to be around in the current world. In other words, the difference between something being attested once and something being unattested is not a very important observation that has to be captured by a theory about human language. Taken to its extreme, this observation suggests that nothing is impossible for human language (as there is no lower boundary of possibility), though many things are highly improbable.

not necessarily the only factor involved. There is also a contrary argumentation placing choric *we* closer to inclusive pronouns, based on the fact that situation in which choric *we* is used are emphasizing collective identity. Such situations are very similar to the in-group usage of inclusive, like in phrases as ‘we linguists’.

¹⁷ See Cysouw (2005d) for a more extensive summary of the attested syncretisms, which also includes dual marking and honorific usage.

Figure 2. Frequency of syncretisms (data extracted from Cysouw 2003b).



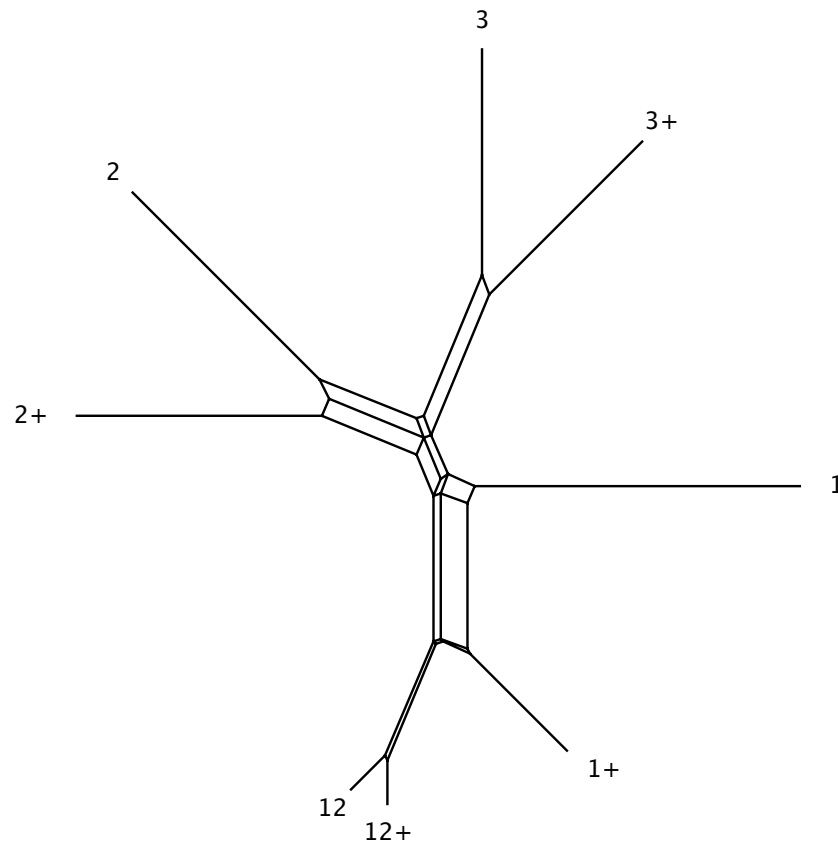
The circumstance that some syncretisms are widespread is a much more robust observation. Exactly how frequent any of the widespread syncretisms is will again strongly depend on the details of the languages sampled. However, I expect that common structures will always be relatively common, independent of the details of the sample of languages. The difference between common and rare is a much more interesting observation to be described by a theory of human language than the difference between attested and unattested. Yet, this perspective immediately leads to the next observations that arises from the distribution in Figure 2: where is the division between common and rare syncretisms to be located? The attested frequencies show a continuous cline between the more common and the less common structures, and nowhere is there a ‘natural’ break between the two. A sharp division between common and rare does not seem to be warranted by the attested variation. This suggests that any theory about person/number marking should be formulated as to capture a continuous cline of variation, and not to separate human languages into two groups (neither “attested vs. unattested” nor “common vs. rare”).

From a perspective claiming that nothing is impossible in human language, but some structures are more likely than others, a cross-linguistic theory of person/number marking will have to be probabilistic in nature.¹⁸ Various kinds of approaches to model and visualize the probabilities of syncretisms have been explored in Cysouw (2007a). Yet another visualization is shown in Figure 3. This figure shows a so-called SPLITS GRAPH of the eight person/number categories as described in Table 4.¹⁹ The basis for this figures are the frequencies of syncretisms for every pair of the eight categories. The more often two categories are attested being expressed by one morpheme in the world's languages, the closer these two categories will end up in the splits graph. The length of the branches indicates how strong the evidence is for each particular grouping (or 'split'). Alternative groupings are shown crossing each other, leading to the rectangles in the graph. The length of the sides of the rectangles is proportional to the frequency of occurrence. For example, there is a rectangle at the bottom of Figure 3, which at the long side separates the first person plural (12, 12+ and 1+) from the other categories. The other shorter side separates first person singular plus exclusive (1 and 1+) from the rest. Both are attested groupings, but the former is clearly more widespread than the latter. Likewise, the length of the individual lines to the leaves of the tree is proportional to the frequency of non-syncretic occurrence of these categories. The category '1' has the longest individual line, and the categories '12' and '12+' the shortest. This indicates that the category 'speaker' is most often expressed with morphological marking of its own. In contrast, separate expressions for the minimal inclusive (12) and augmented inclusive (12+) are much less common.

¹⁸ As remarked by an anonymous reviewer, gradient phenomena do not necessarily imply a gradient theory. There might still be absolute rules governing observed continua, though these have to be embedded into a theory that allows the absolute rules to be turned into gradient effects.

¹⁹ The notion of a splits graph was originally proposed by Bandelt & Dress (1992) and further developed by Dress & Huson (2004). The particular splits graph shown in Figure 3 is a neighborNet made by the program SplitsTree (Huson & Bryant 2006), which draws a splits graph based on the Neighbour Joining tree-building algorithm first proposed by Saitou & Nei (1987). See Bryant et al. (2005) and Cysouw (2007b) for an introduction to this graphical representation with example from linguistics.

Figure 3. Splits graph of person/number categories, based on frequency of morphemic syncretisms (the more frequent the syncretisms, the closer the categories).



A graphical display like the splits graph in Figure 3 is not really a theory about human language, but more of a model of attested frequencies. Such a display might be useful to easily appreciate the relative ubiquity of particular syncretisms, possibly leading to a better understanding of the phenomena to be explained. However, whatever cross-linguistic theory about the structure of human language will be developed on such a basis, it will have to be able to cope with gradients of probability and contrary tendencies. Simply proclaiming that the most commonly attested variant is the norm, and analyzing all remaining variation as structures that are in some way diluted, mixed, or historically corrupted amounts to theoretical prescriptivism. In the righteous appreciation of variation lies the real beauty of any theoretical approach to human language.

5. Conclusion

It is a laudable undertaking to base any theory of person/number marking (or any marking of human language for that matter) on a strong empirical basis of attested variation among the world's languages. Harley & Ritter's (2002b, 2002a) proposals start out in this direction, though unfortunately they do not follow the consequences of the attested variation to its logical conclusion.

Probably every fine looking generalization will at some point run against the counterexamples among the world's linguistic diversity. Yet, incidental exceptions should not necessarily be considered a nuisance for the theory because individual languages simply do not care for world-wide tendencies. In contrast, strong differences in ubiquity of structures should be the foundation of any generalization. To separate the common from the rare, preferably in a continuous manner, is the basic objective of a cross-linguistic theory. A cross-linguistic theory does not necessarily have to capture each and every detail of every human language, but at least the broad preferences should be accounted for. The more detail of the variation is explained, the better.

This sharply distinguishes a cross-linguistic theory from a language-specific explanation. To explain the structures as attested in any individual language, a retelling of the serendipitous historical event leading to the current situation suffices. Of course, more widespread tendencies should be preferred over ad-hoc stories as explanations of the structures in individual languages. However, the quirky details of individual languages should not be forced into spurious generalization. Quite to the contrary, accounting for the large patterns is the real test for a general theory.

Acknowledgments

I thank two anonymous reviewers and the editors for their comments and insistence to make this paper suitable for publication.

References

- Baerman, M. (2002). *The Surrey Person Syncretism Database*.
- Baerman, M. (2004). Directionality and (un)natural classes in syncretism. *Language* 80(4), 807-827.
- Baerman, M. (2005). Typology and the formal modelling of syncretism. In Booij, G., & van Marle, J. (Ed.), *Yearbook of Morphology 2004* (pp. 41-72). Dordrecht: Springer.
- Baerman, M., & Brown, D. (2005). Syncretisms in verbal person/number marking.

- In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 122-125). Oxford: Oxford University Press.
- Baerman, M., Brown, D., & Corbett, G. G. (2005). *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
- Bandelt, H.-J., & Dress, A. W. M. (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* 1(3), 242-252.
- Benincà, P., & Poletto, C. (2005). The third dimension of person features. In Cornips, L., & Corrigan, K. P. (Ed.), *Syntax and Variation: Reconciling the Biological and the Social* (pp. 265-299). Amsterdam: Benjamins.
- Bhat, D. N. S. (2004). *Pronouns*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Bhat, D. N. S. (2005). Third-person pronouns and demonstratives. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 178-181). Oxford: Oxford University Press.
- Boas, F. (1911). Introduction. In Boas, F. (Ed.), *Handbook of American Indian Languages* (pp. 1-83). Washington, DC: Bureau of American Ethnology.
- Bobaljik, J., David. (2008). Missing persons: A case study in morphological universals. *The Linguistic Review* 25(1-2), 203.
- Bolinger, D. (1950). Rime, Assonance, and Morpheme Analysis. *Word* 6, 117-136.
- Borgman, D. M. (1990). Sanuma. In Derbyshire, D. C., & Pullum, G. K. (Ed.), *Handbook of Amazonian Languages* (pp. 15-248). Berlin: Mouton de Gruyter.
- Bryant, D., & Huson, D. H. (2005). Application of Phylogenetic Networks in Evolutionary Studies. *Phylogenetic Networks* 23(2), 254 - 267.
- Buchler, I. R., & Freeze, R. (1966). The Distinctive Features of Pronominal Systems. *Anthropological Linguistics* 8(8), 78-105.
- Chlenova, S. F. (1973). Kategorija chisla v lichnyx mestoimenijax [Category of number within personal pronouns]. In (Ed.), *Lingvotipologicheskie Issledovanija* (pp. 164-201). Moskva: Izdatel'stvo MGU.
- Conklin, H. C. (1962). Lexicographical Treatment of Folk Taxonomies. In Householder, F. W., & Saporta, S. (Ed.), *Problems in Lexicography* (pp. 119-142). Bloomington, IN: Indiana University Press.
- Cowper, E., & Hall, D. C. (2005). The pieces of pi. In Junker, M.-O., McGinnis, M., &

- Roberge, Y. (Ed.), *Proceedings of the 2004 CLA Annual Conference* (pp. Cysouw, M. (2001a). The (a)symmetry of inflectional person marking. Paper presented at ALT IV: Santa Barbara, Calif.
- Cysouw, M. (2001b). The paradigmatic structure of person marking. Radboud University Nijmegen.
- Cysouw, M. (2003a). Against implicational universals. *Linguistic Typology* 7(1), 89-101.
- Cysouw, M. (2003b). *The Paradigmatic Structure of Person Marking*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Cysouw, M. (2005a). A typology of honorific uses of clusivity. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 213-230). Amsterdam: Benjamins.
- Cysouw, M. (2005b). Inclusive/exclusive forms of 'we'. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 162-169). Oxford: Oxford University Press.
- Cysouw, M. (2005c). Syncretisms involving clusivity. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 73-111). Amsterdam: Benjamins.
- Cysouw, M. (2005d). What it means to be rare: the case of person marking. In Frajzyngier, Z., Hodges, A., & Rood, D. S. (Ed.), *Linguistic Diversity and Language Theories* (pp. 235-258). Amsterdam: Benjamins.
- Cysouw, M. (2007a). Building semantic maps: the case of person marking. In Wälchli, B., & Miestamo, M. (Ed.), *New Challenges in Typology* (pp. 225-248). Berlin: Mouton de Gruyter.
- Cysouw, M. (2007b). New approaches to cluster analysis of typological indices. In Köhler, R., & Grzбек, P. (Ed.), *Exact Methods in the Study of Language and Text* (pp. 61-76). Berlin: Mouton de Gruyter.
- Cysouw, M. (2009). The asymmetry of affixation. *Snippets* 20(3), 10-14.
- Cysouw, M. (2010). Semantic maps as metrics on meaning. *Linguistic Discovery*
- Dalrymple, M., & Kaplan, R. M. (2000). Feature indeterminacy and feature resolution. *Language* 76(4), 759-798.
- Daniel, M. (2005). Plurality in independent personal pronouns. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 146-149). Oxford: Oxford University Press.

- Daniel, M., & Moravcsik, E. A. (2005). The Associative Plural. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 150-153). Oxford: Oxford University Press.
- Dobrushina, N., & Goussev, V. (2005). Inclusive imperative. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 179-211). Amsterdam: Benjamins.
- Dress, A. W. M., & Huson, D. H. (2004). Constructing Splits Graphs. *IEEE Transactions on Computational Biology And Bioinformatics* 1(3), 109 - 115.
- Filimonova, E. (2005a). Clusivity cross-linguistically: Common trends and possible patterns. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 399-424). Amsterdam: Benjamins.
- Filimonova, E., ed. (2005b). *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction*. Typological Studies in Language. Amsterdam: Benjamins.
- Fischer, S. (2006). Zur Morphologie der deutschen Personalpronomina – eine Spaltungsanalyse. In Müller, G., & Trommer, J. (Ed.), *Subanalysis of Argument Encoding in Distributed Morphology* (pp. 77-101). Leipzig: Institut für Linguistik, Universität Leipzig.
- Forchheimer, P. (1953). *The Category of Person in Language*. Berlin: De Gruyter.
- Frampton, J. (2002). Syncretism, impoverishment, and the structure of person features. *Chicago Linguistic Society* 38(1), 207-222.
- Gehling, T. (2004). *'Ich', 'du' und andere: Eine sprachtypologische Studie zu den grammatischen Kategorien 'Person' und 'Numerus'*. Sprache, Kommunikation, Wirklichkeit: Linguistische Studien zur Semantik und Pragmatik. Münster: LIT.
- Gehling, T. (2006). Die Suche nach dem anderen ihr. Zur Inklusiv-Exklusiv-Distinktion in der Zweiten Person. In Gehling, T., Voss, V., & Wohlgenut, J. (Ed.), *Einblicke in Sprache* (pp. 153-180). Berlin: Logos.
- Georgi, D. (2006). A distributed morphology approach to argument encoding in Kambera. In Müller, G., & Trommer, J. (Ed.), *Subanalysis of Argument Encoding in Distributed Morphology* (pp. 1-22). Leipzig: Institut für Linguistik, Universität Leipzig.
- Goffman, E. (1979). Footing. *Semiotica* 25(1-2), 1-29.
- Hanson, R. (2000). Pronoun acquisition and the morphological feature geometry. *Calgary Working Papers in Linguistics* 22, 1-14.
- Harley, H., & Ritter, E. (2002a). Person and number in pronouns: a feature-

- geometric analysis. *Language* 78(3), 483-526.
- Harley, H., & Ritter, E. (2002b). Structuring the bundle: A universal morphosyntactic feature geometry. In Simon, H., & Wiese, H. (Ed.), *Pronouns - Grammar and Representation* (pp. 23-39). Amsterdam: Benjamins.
- Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1), 119-132.
- Helmbrecht, J. (2004a). Ikonizität in Personalpronomina. *Zeitschrift für Sprachwissenschaft* 23(2), 211-244.
- Helmbrecht, J. (2004b). Personal Pronouns: Form, Function, and Grammaticalization. Erfurt.
- Hollenbach, B. E. (1970). Inclusive Plural: A Further Look. *Linguistics* 60, 27-32.
- Holt, D. (1999). *Pech (Paya)*. Languages of the World/Materials. Munich: Lincom.
- Hopkins, E. B. (1986). Pronouns and pronoun fusion in Yaouré. In Wiesemann, U. (Ed.), *Pronominal Systems* (pp. 191-204). Tübingen: Narr.
- Howe, S. (1996). *The Personal Pronouns in the Germanic Languages: A study of personal pronoun morphology and change in the Germanic languages from the first records to the present day*. Studia Linguistica Germanica. Berlin: De Gruyter.
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2), 254-267.
- Jespersen, O. (1924). *The Philosophy of Grammar*. London: Allen & Unwin.
- Laycock, D. (1977). Me and you versus the rest. *Irian* 6, 33-41.
- McGinnis, M. (2005). On markedness asymmetries in person and number. *Language* 81(3), 699-718.
- McGregor, W. B. (1990). *A Functional Grammar of Gooniyandi*. Studies in Language Companion Series. Amsterdam: Benjamins.
- McGregor, W. B. (1996). The pronominal system of Gooniyandi and Bunuba. In McGregor, W. B. (Ed.), *Studies in Kimberley Languages in Honour of Howard Coate* (pp. 159-173). Munich: Lincom.
- McKay, G. R. (1978). Pronominal person and number categories in Rembarrnga and Djeebbana. *Oceanic Linguistics* 17, 27-37.
- Moravcsik, E. A. (1994). Group plural – associative plural or cohort plural. *Linguist List* 5(681),
- Moravcsik, E. A. (2004). A semantic analysis of associative plurals. *Studies in Language* 27(3), 469-503.

- Müller, G. (2006a). Notes on paradigm economy. In Müller, G., & Trommer, J. (Ed.), *Subanalysis of Argument Encoding in Distributed Morphology* (pp. 161-195). Leipzig: Institut für Linguistik, Universität Leipzig.
- Müller, G. (2006b). Subanalyse verbaler Flexionsmarker. In Breindl, E., Gunkel, L., & Strecker, B. (Ed.), *Grammatische Untersuchungen, Analysen und Reflexionen. Gisela Zifonun zum 60. Geburtstag* (pp. 183-203). Tübingen: Narr.
- Nevins, A. I. (2003). Do person/number syncretisms refer to negative values? Paper presented at 77th meeting of the LSA: Atlanta.
- Palácio, A. P. (1986). Aspects of the morphology of Guató. In Elson, B. F. (Ed.), *Language in global perspective: Papers in honor of the 50th anniversary of the Summer Institute of Linguistics 1935-1985* (pp. 363-372). Dallas, Tx.: Summer Institute of Linguistics.
- Payne, D. L., & Payne, T. E. (1990). Yagua. In Derbyshire, D. C., & Pullum, G. K. (Ed.), *Handbook of Amazonian Languages* (pp. 249-474). Berlin: Mouton de Gruyter.
- Pike, K. L. (1965). Non-linear order and anti-redundancy in German morphological matrices. *Zeitschrift für Mundartforschung* 32 (3/4), 193-221.
- Ritter, E., & Harley, H. (1998). Meaning in morphology: a feature-geometric analysis of person and number. Paper presented at 40th GLOW Colloquium: Tilburg.
- Rumsey, A. (1996). On some relationships among person, number, and mode in Bunuba. In McGregor, W. B. (Ed.), *Studies in Kimberley languages in honour of Howard Coate* (pp. 139-148). Munich: Lincom.
- Rumsey, A. (2000). Bunuba. In Dixon, R. M. W., & Blake, B. J. (Ed.), *Handbook of Australian Languages* (pp. 34-152). Oxford: Oxford University Press.
- Saitou, N., & Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4), 406-425.
- Sauerland, U. (2003). A new semantics of number. In Young, R. B., & Zhou, Y. (Ed.), *Proceedings of SALT 13* (pp. 258-275). Cornell, Ithaca: CLC Publications.
- Schwartz, L. J. (1986). The function of free pronouns. In Wiese, U. (Ed.), *Pronominal Systems* (pp. 405-436). Tübingen: Narr.
- Siewierska, A. (2004). *Person*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Siewierska, A. (2005a). Gender distinctions in independent pronouns. In

- Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 182-185). Oxford: Oxford University Press.
- Siewierska, A. (2005b). Third-person zero of verbal person marking. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Ed.), *World Atlas of Language Structures* (pp. 418-421). Oxford: Oxford University Press.
- Siewierska, A., & Bakker, D. (2005). Inclusive and exclusive in free and bound pronouns. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 151-178). Amsterdam: Benjamins.
- Siewierska, A., & Bakker, D. (2006). Bi-directional vs. uni-directional asymmetries in the encoding of semantic distinctions in free and bound person forms. In Nevalainen, T., Klemola, J., & Laitinen, M. (Ed.), *Types of variation: diachronic, dialectical and typological interfaces* (pp. 21-50). Amsterdam: Benjamins.
- Simon, H. (2005). Only you? Philological investigations into the alleged inclusive-exclusive distinction in the second-person plural. In Filimonova, E. (Ed.), *Clusivity: Typology and Case Studies of the Inclusive-Exclusive Distinction* (pp. 113-150). Amsterdam: Benjamins.
- Sokolovskaya, N. K. (1980). Nekotorye semanticeskie universalii v sisteme licnyx mestoimenij [Some semantic universals in systems of personal pronouns]. In Vardul', I. F. (Ed.), *Teorija i tipologija mestoimenij [Theory and typology of pronouns]* (pp. 84-102). Moskva: Nauka.
- Swiggers, P., & Wouters, A. (1998). *The Tékhne Grammatiké van Dionysius Thrax: De oudste Spraakkunst in het Westen*. Orbis Linguarum. Leuven: Peeters.
- Thomas, D. D. (1955). Three analyses of the Ilocano pronoun system. *Word* 11(2), 204-208.
- Tuite, K. (1997). *Svan*. Languages of the World/Materials. Munich: Lincom.
- Uhlig, G. (1883). *Dionysii Thracis Ars Grammatici*. Grammatici Graeci. Leipzig: Teubner.