

Syntagmatical Variation in the World's Pronominal Systems

Michael Cysouw
General Linguistics and Dialectology
University of Nijmegen
m.cysouw@let.kun.nl

Abstract

The variation in pronominal marking in the world's languages is often underestimated. This paper categorises the diversity of the syntagmatical variation of pronominal elements across languages. The syntagmatical variation is reduced to a few independent continua: bound-free, frequent-infrequent, meaning-setting. Due to linguistic structures that happen to be common in the languages that are studied in the Occidental tradition, there are also some overestimations of the variability of pronominal marking. Some linguistic structures that may seem normal from a West European point of view, are extremely marked from a worldwide view. Three non-universals are formulated to highlight the cross-linguistic rareness of some aspects of syntagmatical marking of pronominal elements.

Introduction

Pronominal reference, words like ‘I’ and ‘you’, are widely claimed universal elements of human languages (cf. Wierzbicka 1996:36-38). Whether true or not, pronominal reference is at least an object of study that provides lots of data on different languages, a promising perspective for a cross-linguistic study. Strange enough, there are hardly any published studies of pronominal marking comparing a wide variety of the world’s languages. The only works with a true cross-linguistic perspective that I am aware of are Forchheimer (1953), Ingram (1978) and Mühlhäusler et al. (1990). This scarcity indicates that although a lot is claimed on pronominal marking, the cross linguistic test of those theoretical claims is still in its infancy.

In this paper I will take a step towards a theory on pronominal marking that takes into account the variation that is found in the rather diverse languages still spoken in the world. The basic goal of this paper is to describe a few dimensions of variation with respect to pronominal marking. It is this variation that has to be the bottom line of every claim about universality of pronominal marking. If the *known* variation of human language is not accounted for, any statement about the *possible* variation is still far from home. Secondly I will propose a few claims on the general pattern of pronominal marking across languages.

Goal and method

In recent years most grammatical research focused on universal properties of human language. Central to this inquiry is to find restrictions on the possible structures of human language. One way to deal with this goal is by way of *typology*. In typology a representative sample of the world’s languages is categorised in a few different *language types*. Hopefully, some types that a priori seem feasible as possible structures for a human language, do not occur in the sample – or are only rarely found. Such unusual types indicate restrictions on the possible structures of human language.¹

A basic problem for the typological method is to find a way to actually compare the wide variety of structures that can be found in the languages of the world. To be able to compare, a basis for comparison has to be defined. In the case of linguistic structure, such a basis of comparison often falls short of capturing the variety of the structures languages use. Stated differently, the problem is how to define types for the typology that capture the wide variation of occurring linguistic structures. The infamous typology of basic word order, for instances, defines six different types of languages, based on the possible order of verb, subject and object in a transitive sentence. This typology though is ‘the wrong question’ for those languages that have so-called ‘free’ word order, like Acehnese (Austronesian, Indonesia).² In

‘free’ word order languages word order is not at all free, as there are normally strong contextual and pragmatic constraints on word orders. However, there is no real *basic* word order, in the same sense as English is SVO. It is the wrong question to ask to which of the six word order types Acehnese belongs (Durie 1985:191-193). This shows that the six word order types do not classify the complete variation of linguistic structures found in the world’s languages as there are languages that do not belong to any of the six types.

To find suitable definitions of types for a typology is a difficult task. The prominent view of linguistic structure is still strongly based on classical analyses of Latin and a few other European languages. The languages of Europe, though, represent only a tiny fraction of the world’s linguistic diversity, and, as a consequence, questions formulated from this background are often heavily biased towards structures found in European languages. The goal of the research presented here is to test linguistic analyses on such European bias, and work towards cross-linguistic feasible categorisation of linguistic structure. I call this kind of research ‘cross-linguistic’. The ultimate result of a cross-linguistic study is a classification that can be used to make a typology.

In the present study, *pronominal systems* are the object of cross-linguistic scrutiny. There is a rather widespread unexplicated assumption in grammatical theory that every language has one, and only one, principal pronominal system. This ‘principal’ system to mark person-deixis is, for instance, by way of independent pronouns in the Germanic languages, but by way of verbal inflection in most Amerindian languages. If other person-deixis morphemes exist in a language, these will be assigned secondary status – they are e.g. ‘agreement’ or ‘emphatic pronouns’. For this cross-linguistic research into the diversity of pronominal systems in the world’s languages, I refrain from that assumption, treating all person-deixis morphemes in a particular language on an equal basis. This implies that a language may have – and most indeed have – more than one pronominal system.

Pronominal systems are built from pronominal elements. Cross-linguistically, it is not immediately obvious how to define these notions. Basically, pronominal elements are *linguistic signs for person deixis*, and pronominal systems are paradigms of pronominal elements. An important addition to these basic definitions is that the linguistic signs should be *specialised* for person deixis. In Southeast Asian languages, for instance, it is quite normal to use full nouns for person deixis (Cooke 1968). In English this is only possible in highly marked contexts, like when a mother speaking to her child says ‘mommy is going upstairs’, referring to herself with the noun ‘mommy’. These full nouns are not specialised instruments for person marking, and are not included as pronominal marking in this study. This implies, consequently, that there are languages, like Thai (Cooke 1986:6-70), that do not have pronominal marking as a grammatical category. Thai has person deixis – every language has – but it does not have a specialised grammatical category to mark that deixis.

Besides this formal criterium there are a few functional/semantic constraints on the data to be included in this study. As for ‘deixis’, demonstrative deixis is excluded. Demonstratives are a sort of deixis, but not basically related to person. On top of that, demonstratives seem to be a rather neatly restricted domain of linguistic marking in the languages of the world, and deserve separate attention. Other forms of deixis which I exclude from the present discussion are reflexives and reciprocals. Just as with demonstratives, these elements seem to form a class of themselves, although an incidental overlap with ‘basic’ person deixis is found. Finally I will not deal with pronominal possession, neither with other adnominal person marking, e.g. construction like ‘me, the king’. As well demonstratives, reflexives, reciprocals as pronominal possession ask for an individual cross-linguistic survey, I will not treat them as secondary phenomena to person deixis.

Keeping these constraints in mind, I take *speaker reference* as the prototypical form of person deixis. All morphemes that refer to the speaker are included in the database as person deixis morphemes. For Dutch this includes ‘ik’, ‘mij’ and the zero inflection of the verb. To these ‘first person’ elements all elements that belong to the same paradigm are then added. This forms the complete set of pronominal elements that are included in the present study.

To bring order to the overwhelming diversity of these pronominal elements, I distinguish two main classificatory dimensions: syntagmatical and paradigmatic variation. The syntagmatical dimension classifies pronominal elements with respect to their interaction with other morphemes in the language – characteristics such as ‘bound’, ‘case’ etc. The morphemes with identical syntagmatical status make up a ‘pronominal system’. The paradigmatic dimension then classifies the variation within each pronominal system. The present paper will deal with the syntagmatical dimension only.³

The data for this paper are compiled from published sources only. This means that I have to trust the descriptions that are made in the literature. To prevent too much influence from ideosyncratic analyses in a particular grammar, I have taken the morphophonological analysis of a language as basis. I assume the morphophonological analyses to be comparable in their insights, irrespectable of the different morphosyntactic perspective a description may take. Consequently, I interpret the morphophonological form fairly strict, meaning practically that a category does not exist if there is no morpheme to code for it in a language.

For instance, zeros do not exist, unless there is a strong morphophonological reason. Methodologically, proposing a linguistic category predicts a language should have specialised linguistic means to express that category. What should be done if there turns out to be a language that does not express that category by any specially ‘designed’ element? In a strict Popperian way of falsification one would have to search for a new theory, but this step is almost never taken. The obvious solution is to temporarily fill the gap in the theory with some kind of patch, an ad hoc solution to be able to go on with the theory. To use such a patch is an accepted practice in

science, but it still is a sign of weakness for the theory. To proclaim a certain category to be zero in a language is such a patch, a weakness in the descriptive paradigm. The more zeros needed, the weaker the theory. If zeros keep popping up, the theory and the categories of this theory should be reconsidered.

I will refrain here from quantitative analyses of the occurrence of linguistic structures. There are too many factors that obscure the interpretation of statements like ‘in 70 languages out of a sample of 100’. Such skewing in a cross-linguistic sample could be due to linguistic properties, but also to unknown family ties or large-areal bias in the languages of the sample. Instead of a quantitative approach, I have chosen to restrict myself to qualitative analyses.

First, I will deal with a description of the wide variety of syntagmatical structures of pronominal systems. The next paragraph will describe a few dimensions of syntagmatical variation which have to be taken into account for a description of the actual cross-linguistic variation. Certain kinds of variation are widely known and discussed (like ‘clitic’ versus ‘free’ pronouns), other dimensions of variation are hardly acknowledged. Second, I will propose three *universals*, three restrictions on the linguistic structures in the sample. I have until now looked in detail at roughly 250 languages and numerous general accounts on families or other groupings of languages. The proposed universals are near universal, i.e. there are only a few counterexamples that I know of. All the counterexamples will be discussed here. The three universals are formulated negatively; they tell something about what is not – or hardly – attested in the world’s languages. For that reason I call them *non-universals*.

Syntagmatical variation

To bring some order to the overwhelming diversity of person marking in the world’s languages, I distinguish between two different dimensions of variation, paradigmatic and syntagmatic variation. Paradigmatic variation is the variation within one paradigm, i.e. variability of the group of elements that have roughly the same place in the systematicity of a language. The subject of this section though is the syntagmatic variation. Syntagmatic variation is the variation in the *relation* between the pronominal elements and other elements in the language. It is, for example, the case-relation between a pronoun and a verb that distinguishes the subject pronoun in English from the object pronoun. Other languages distinguish between different pronominal paradigms on other grounds.

Paradigmatic variation is, so to say, the internal variation of pronominal systems. Syntagmatic variation, oppositely, is the variation between different pronominal systems within the systematicity of a language. These two dimensions of variation, syntagmatic and paradigmatic, are independent aspects for a typology. There is, a priori, no reason to assume that there are restrictions on the occurrence of any

combination of these dimensions. Whether there are any correlations is a subject for further research.

I distinguish between two aspects of syntagmatical variation. There is variation in the *form* of the relation between pronominal elements and other linguistic elements, and in the *content* of the relation between person deixis and other linguistic elements. The *form* of the relation refers, for instance, to aspects like ‘bound’ marking or ‘free’ marking. There is much more to this dimension than only these two well known categories. This variation will be discussed in the next subsection.

The dimension of the syntagmatical *content* of the relation shows an even more extensive variation, to be tackled in the second subsection.⁴ This dimension confronts variation like the relation of the pronominal system to the *meaning* of the predicate, i.e. variation in case or theta-roles. On the other side it also confronts variation in the relation of the pronominal system to the *setting* of the predicate, i.e. variation in tense, aspect or mood marking. This variation will be discussed in the second subsection. Just like paradigmatic and syntagmatical variation are a priori independent parameters, in the same sense the two subdimension of syntagmatical variation, form and content, are a priori independent from each other.

Form: bound vs. free, and the like

The opposition between *bound* and *free* marking is a well-known syntagmatical distinction. The English pronouns are free (‘I’, ‘you’, etc.), the pronominal suffixes in Spanish are bound (‘cant-**o**’, ‘cant-**as**’, etc.). Also, there exist an extensive literature on intermediate cases between bound and free, called *clitics*: phenomena like the French object-pronouns (‘je l’ai vu’). The boundary between free, clitic and bound marking is often hard to pin down, as it is more of a continuum than a threefold typology. I will not discuss this continuum further, as I want to draw attention here to some other aspects of syntagmatical variation of pronominal marking that are not widely acknowledged. For the rest of this paper, I’ll use the terms *bound* and *free* to designate the prototypical extremes of a continuum.

Not all bound marking is just bound marking; bound marking exists in different guises. The status of bound marking in the grammar of a particular language depends on the grammatical characteristics of the *root* onto which the pronominal marking is bound. Languages vary for instance in the amount of roots that allow for bound pronominal marking. One extreme of this dimension is found in Kalam (East New Guinea Highlands, Papua New Guinea). Kalam has a *closed* class of pronominally inflected verbs: there are roughly a 100 of these ‘finite verbs’, and – stressing the extremity of this case – 15 out of those 100 are responsible for 89% of the appearances of verbs in texts (Pawley 1993:87). This is kind of an extreme case of the well-known phenomenon that in some languages a small group of auxiliaries are used excessively. The other extreme of this dimension is found in Nootka (Waka-

shan, Canada) where almost all lexical roots in the language can function as predicate, and are, in that function, inflected for person (Rose et al. 1984:1).⁵ The European type of language, with an open class of verbs that take bound pronominal marking, is an intermediate case in this continuum.

Not only the *amount* of roots shows variation, also the *frequency of occurrence* of pronominally inflected roots can differ. A language may have an open class of pronominally inflected roots, but uses these only sparsely. This occurs for instance in languages that have a special ‘serialisation’-form for predicates. In Siroi (Madang, Papua New Guinea) there are other verb forms, besides inflected verbs, that are used regularly. One of them is the typical New Guinean ‘dependent’ form that can be used in long series as shown in (1), with only one pronominally inflected verb at the end. This is a rather different type of person marking from a language where all verb forms have to be inflected for person.

- (1) k-umba kule pis-mba mwi minyang-a mb-umba
 go-**Dep** water bathe-**Dep** hand wash-**Dep** ascend-**Dep**
- ne marasin ti-n-i
 2SgPron medicine give-2Sg-**1Sg**

‘I will go and bathe in the river, wash (my) hands, come up and
 give you medicine’ (Wells 1979:79)

Variation in the frequency of occurrence is also found with free pronominal marking. However, free pronouns are, by definition, not bound to the occurrence of other morphemes, and this makes it difficult to compare frequencies. The variation in occurrence of free pronouns has to be compared relative to sentential structures. Some languages make a much more structural use of their pronouns in a sentence compared to other languages. The infamous ‘pro-drop’ parameter confronts this variation. Put bluntly, the pro-drop parameter distinguishes two types, the ‘non-pro-drop’ type like English, where the pronoun is obligatorily present, and the ‘pro-drop’ type like Spanish, where the pronoun is optionally present. The actual variation though is much larger as this dichotomy might suggest.

In ‘pro-drop’ languages the frequency of ‘drop’ varies. It is difficult to pinpoint the differences, but controlled cross-linguistic counting of frequencies yields rather different amounts of ‘drop’ for different languages (cf. the studies of individual languages in Givón 1983). Also the non-pro-drop languages show variation. In non-pro-drop languages every sentence has to show its arguments overtly, they are not allowed to be dropped. Differences occur because the structure of the *sentence* can change quite radically cross-linguistically. Supyire (Gur, Mali) is non-pro-drop, in a sense, as there has to be a subject pronoun if there is no nominal subject present (Carlson 1990:426 e.v.). As long as the subject remains the same, there can occur long stretches of text in Supyire that are syntactically one single sentence, and that

consequently only need one single subject pronoun. Example (2) is the translation of one singly sentence form Supyire. Only the first pronoun ‘she’ is actually present in the original (Carlson 1990:1024-1025). Supyire is non-pro-drop just like English, but the actual marking of person differs strongly.

- (2) Then **she** cooked that mush and ate till she was stuffed, and went and gave the men’s (food to them) and finished (giving out the food) and went and got the dishes and came and stretched herself and said...

Note that in both the example (1) from Siroi and (2) from Supyire are not that ‘exotic’ when compared to the English translations. Both the occurrence of pronominally inflected verbs, as the occurrence of the pronoun, can be strongly reduced in English. However, to put the utterances (1) and (2) into one single running sentence, as is done here to give a close translation of the examples, is a little awkward in English. Most speakers of English would opt for different constructions in these cases, raising the amount of overt pronominal morphemes.

Content: meaning vs. setting

A complete different dimension, independent of the *form* of the syntagmatical relation, is the variation in the *content* of the syntagmatical relation. A lot of languages have different pronominal elements that are distinguished because they have a different function in the systematicity of a language. A well known example of this variation is the opposition between subject (‘I’, ‘he’, etc.) and object (‘me’, ‘him’, etc.) pronouns in English. The difference between these two sets of pronouns is the content of their relation to the sentence predicate. In this section I will give a rough sketch of the possible oppositions languages make in this respect.

Table 1: Grebo combined mood markers/pronouns

| | <i>Indicative</i> | | <i>Subjunctive</i> | | <i>Conditional</i> | | <i>Result</i> | | <i>Emphatic</i> | |
|--------|-------------------|------------------|---------------------|-----------------|---------------------|---------------------|---------------------|--------------------|-------------------|-------------------|
| | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> |
| 1 | ne ¹ | a ^{3/4} | bè ^{3-2/4} | ba ³ | ne ^{1/2} | ba ^{4-1/4} | ne ^{1/2} | a ^{4-3/4} | mo ¹ | amo ³³ |
| 2 | ∅ | a ² | bè ³ | ba ² | ne ^{4-1/4} | ba ^{1/2} | ne ^{4-3/4} | a ^{1/2} | mo ^{2/3} | amo ¹³ |
| 3 High | | o ² | | bo ² | | bo ^{1/2} | | o ^{1/2} | | no ² |
| 3 Low | | e ² | | be ² | | be ^{1/2} | | e ^{1/2} | | ne ² |

(Innes 1966: 50-51)

The first variation I will highlight is variation in tense, aspect or mood marking. An example of this is found in Grebo (Kru, Liberia). Grebo has free pronouns in differ-

ent guises that mark mood-differences (see Table 1). In this way, the pronoun actually adds something to the meaning of the utterance, besides marking the person-deixis. The pronoun specifies the ‘setting’ of the predicate.

The specification of the setting of the predicate can also occur inflectionally on the predicate. Tense, aspect or mood markers can fuse with person deixis to form distinct inflectional paradigms. An example is Yali (Dani-Kwerba, New Guinea). In Yali, the person deixis is combined with past tense suffixes (see Table 2). The specification of the ‘setting’ combined with person deixis thus is independent of the *from* of the marking of person deixes. Both bound and free person deixis can show variation with respect to the marking of the setting.

Table 2: Yali combined past tense/pronominal suffixes

| | <i>Today's past</i> | | <i>Intermediate past</i> | | <i>Far past</i> | |
|---|---------------------|-------------|--------------------------|-------------|-----------------|-------------|
| | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> | <i>Sing</i> | <i>Plur</i> |
| 1 | ...-i | ...-e | ...-ik | ...-uk | ...-ey | ...-ow |
| 2 | ...-en | ...-ep | ...-in | ...-ip | ...-en | ...-ep |
| 3 | ...-ek/a | ...-esa | ...-si | ...-sa | ...-eb | |

(Fahner 1979:74-76)

Another possibility of syntagmatical variation in human language is that the *meaning* of the predicate itself controls different forms of person deixis. This is mostly called variation in ‘case’ or ‘theta roles’. As ‘case’ or ‘theta’ marking is widely known, and I will here not linger longer on this topic. Only note that, here again, the *form* of the marking is independent of possibility to mark case variations. Case can as well be marked on free pronouns as inflectionally. There seem to be some restrictions on inflectionally marked case roles. This will be discussed below in connection to non-universal 3.

These two possibilities for syntagmatical variation are not discretely distinguished; they form a continuum. On one end of the continuum, the pronominal variation *adds* meaning, like in the case of tense marking in Yali, on the other end the pronominal variation is *controlled by* the meaning of the predicate. The example of Grebo, where the pronouns mark mood is already an intermediate case, as mood marking is normally restricted by the meaning of the predicate.

As a real intermediate case I will present here the so called ‘fluid’ case marking (Dixon 1994:78-83). An example of this phenomenon is found in Northern Pomo (Pomoan, USA). In Northern Pomo there are two different paradigms of pronouns, one ‘agent-like’ and the other ‘patient-like’. With a lot of predicates the choice for

either of these is controlled by the meaning of the predicate, as for instance in (3). A verb like ‘to jump’ needs a ‘agent-like’ pronoun, but a verb like ‘to be sick’ needs a ‘patient-like’ pronoun.

- (3) a. ?a· p^hdíw
 1SgAgt jump
 ‘I jumped’ (Mithun 1991:518)
- b. ʔo· ʔt^hál.
 1SgPat be sick
 ‘I am sick’ (Mithun 1991:518)

With some predicates though, this situation is reversed: the pronoun determines the precise meaning of the predicate. This happens if the predicate allows as well for an ‘active’, as for a more ‘passive’ reading, for instance with the verb ‘to cough’ in (4). Again, this variation can be found marked by free pronouns, as in the examples of Northern Pomo, as well as with inflectionally marked pronominal paradigms.

- (4) a. ?a· k’lú·k’lú·w.
 1SgAgt cough
 ‘I coughed (intentionally)’ (Mithun 1991:520)
- b. ʔo· k’lú·k’lú·w.
 1SgPat cough
 ‘I coughed (involuntarily)’ (Mithun 1991:520)

Concluding, there is a continuum with respect to the *form* of the syntagmatical variation, roughly speaking between free and bound, and a continuum with respect to the *content* of the syntagmatical variation, as shown in Figure 1. These two continua seem to be completely independent parameters of variation.

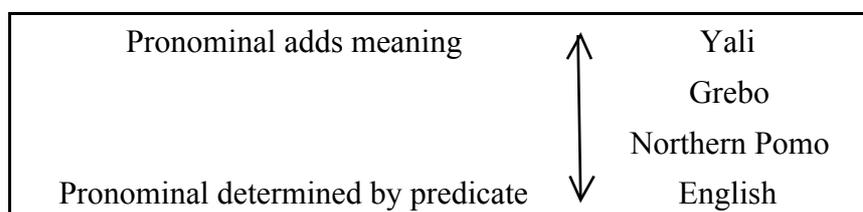


Figure 1: Continuum of content-based variation

Three non-universals

In this section I will propose a few restrictions on the possible syntagmatical variation of human language, consistent with the goal of typological research as put forward in the methodological section. Such restrictions are normally formulated as

universals. Universals in typology should not be confused with universals as put forward in generative syntactic frameworks where ‘universal’ is used in its literal sense, a characteristic that holds for every possible human languages. A ‘universal’ in typology, oppositely, is a technical term for a specific kind of correlation between two characteristics of language in a sample of the world’s languages. These universals normally have counterexamples, and are thus not at all universal in the literal sense. Universals are strong correlations, and, as in every correlation, there are exceptions to it. Typological universals without exceptions are to be distrusted.

Instead of formulating my universals straightforward, I will formulate them as three *non-universals* related to the syntagmatical variation in the world’s languages. Universal in typology are statements of the form ‘If X then Y’. The non-universal I propose, have the general form ‘If X then *Not* Y’. This is not merely a difference in formulation. The characteristic Y, which I claim does *not* occur – under condition X, is a characteristic which is often assumed to be perfectly normal for a language to have. To stress the empirical problems with these widespread assumptions, I have formulated the universals in this section in this negative form. With each non-universal I will discuss the few counterexamples I have found. These counterexamples are the only ones I am aware of after inspecting a few hundred languages. Still this sample represents only a tiny fraction of the 7000, or so, languages in the world still spoken, and I am perfectly aware that there will be more counterexamples. I would be very glad to hear of such.

No paradigmatic equivalence

The first non-universal is about *coreferential* pronominal systems. Often languages have different pronominal systems that can occasionally co-occur with one predicate, but still refer to the same participant. In Dutch, the inflectional pronominal system and the subject pronouns are obligatorily used both, and they refer to the same referent, as shown in (5).

- (5) a. **Ik** loop- \emptyset
 b. **Jij** loop-t

| | 1Sg | 2Sg | 2Hon | 3Msg | 3Nsg | 3Fsg | 3Pl | 2Pl | 1Pl |
|-------------|------------------|-------|------|------|------|------|--------|-----|-----|
| <i>Infl</i> | ...- \emptyset | ...-t | | | | | ...-en | | |
| <i>Free</i> | ik | jij | u | hij | het | zij | jullie | wij | |

Figure 2: Paradigmatical inequivalence in Dutch subjectreference

ent, as in Dutch. In that case the double marking is not redundant, and emphatic double marking adds referential information.

Table 3: Malakmalak free pronouns and object suffixes

| | <i>Free pronouns</i> | | <i>Object suffixes</i> | |
|---------------|----------------------|---------------|------------------------|---------------|
| | <i>Singular</i> | <i>Plural</i> | <i>Singular</i> | <i>Plural</i> |
| <i>1</i> | ŋa | yawot | ...-ārin ^y | ...-yörö |
| <i>2</i> | waŋari | nukut | ...-nunu | ...-nuŋkuŋu |
| <i>1+2</i> | yaŋki | yēŋkit | ...-nuŋku | ...-āŋpuŋu |
| <i>3 Masc</i> | yöntön | wöŋöntön | ...-nö | ...-wöŋö |
| <i>3 Fem</i> | nöntön | | ...-ŋayi | |

(Birk 1976:30-31)

Agreement is highly marked

Things get even worse for agreement. As mentioned in the last section, most languages have a possibility to double the marking of a referent. This double marking is often called ‘agreement’. ‘Agreement’ can also occur between a full noun and a pronominal system, but I am only concerned here with ‘agreement’ between two different pronominal systems. Although most languages *can* double the marking of their pronominal elements, it is highly marked if they do so.

Non-universal 2

If there are two syntagmatically different pronominal systems in a language that can possibly refer to the same referent, then it is highly unusual that both systems co-occur with one predicate.

This universal can be interpreted on two different ways. First *linguistically*, within the structure of most language double marking is highly marked. For instance in a language like Spanish, there is one inflectional pronominal system, suffixed to the verb. Occasionally an independent pronoun can be added for emphasis. The occurrence of suffix *and* free pronoun is marked in Spanish. Second, the non-universal can be interpreted *cross-linguistically*. There are languages that, within the structure of the language, regularly double the marking of the pronominal elements. The examples from Dutch in (5) are the unmarked structure in the language to refer to speaker or addressee. Languages like Dutch that use double marking as the normal structure, are *cross-linguistically* rare.

Outside Europe, there are only a few languages where I have found examples of regular double marking. The first examples to be mentioned here are from two Nilotic languages, Päkot (Nilotic, Kenia) and Lotuho (Nilotic, Sudan). In these languages there is obligatory double marking of referents, both as prefix and as suffix on the verb. An example from Päkot is shown in (7) and an example from Lotuho in (8).

(7) α -ipu:ce-an-ye
 1Sg-wipe-1Sg-Imperfect
 ‘I wipe’ (Rottland 1982:133)

(8) á-ḃwaxa-ni
 1Sg-dig-1Sg
 ‘I am digging’ (Tucker et al. 1966:470)

Interestingly, these two languages are the only two languages in the Nilotic family that show this double marking. All the other languages have maximally one obligatory pronominal system, any coreferential systems are *linguistically* marked. This is reminiscent of the situation in Germanic, where only Dutch and German have regular double marking of pronominals.⁶ The Nilotic situation is even more extreme as Päkot and Lotuho do not belong to the same subbranch of Nilotic and are presently not in geographical contact – Päkot is Southern Nilotic, Lotuho is Eastern Nilotic and the whole Teso-Turkana branch of Nilotic is spoken in between the two. In general it seems that double marking is an exception, even within a group of close relatives.

The other two examples of double marking that I currently know of are from Taba (Austronesian, Indonesia) and Tehuelche (Chon, Argentina). They are both rather awkward cases, but they are mentioned for completeness. In Taba only a small subset of predicates have obligatory double marking: the ‘secretion’ verbs, cf. (9). Although the double marking is regular in these cases, this can be interpreted as a case where double marking is *linguistically* marked, as it is only a small closed class of verbs in this languages that get regular double marking.

(9) k-sio-k
 1Sg-shit-1Sg
 ‘I shit’ (Bowden 1998)

Tehuelche is an almost disappeared language, once spoken in Patagonia in Argentina. Tehuelche has free pronouns that are only used for emphasis. Curiously though, these free pronouns, in the non-singular, consist of two coreferential pronominal elements, as shown in Table 4. They have the same pronominal prefixes as are found on predicates, combined with the singular free pronouns as suffixes.

Double marking of pronominal elements seems to be highly marked in the languages of the world, either linguistically or cross-linguistically. If such double marking is called ‘agreement’, then it should be concluded that this kind of ‘agreement’ hardly occurs.⁷

Table 4: Tehuelche free pronouns

| | <i>Sing</i> | <i>Dual</i> | <i>Plur</i> |
|---|-------------|-------------|-------------|
| 1 | ja: | o-k-wa | o-š-wa: |
| 2 | ma: | m-k-ma: | m-š-ma: |
| 3 | ta: | T-k-ta: | T-š-ta: |

(Fernandez Garay 1993:264)

Triple marking is unusual

The next observation, to be formulated as the third non-universal, has a somewhat different scope. In theories of case marking, or theta roles, there is often a special status proposed for three cases: the so-called ‘core’ cases. These core cases are differently named in the various theories, but all go back to the trinity ‘subject’, ‘object’ and ‘indirect object’. Implicitly, there is a break proposed between these three and any other cases. Looking, though, at the cross-linguistic variation of person-marking, it seems better to propose a break after two core cases.

Part of the languages of the world mark pronominal references to these core cases inflectionally on the verb. Most European languages only mark one core case, the ‘subject’, on the verb. Other languages have more than one inflectional paradigm marking different core cases. But even if there are more than two inflectional paradigms differentiated in the morphology of a language, it is very uncommon for more than *two* to be marked together, overtly on a verb.

Non-Universal 3

If there are more than two inflectional pronominal systems – with different referents – in a language, then it is highly unusual that more than two co-occur with one predicate.

From the analysis of pronominal marking, there is a different status for two participants in an utterance, although the semantic roles they fulfill with respect to the verb can change, even within a language.

Take Chickasaw (Muskogean, USA) as an example. In Chickasaw, as in most Muskogean languages, there are three different paradigms of pronominal affixes. Following (Munro et al. 1982) they are often referred simply as paradigms I, II and

III. The use of the paradigms is roughly indicated by the descriptions ‘agent’, ‘patient’ and ‘dative’, as in Table 5.

Table 5: Chickasaw pronominal affixes

| | | <i>agent</i> | <i>patient</i> | <i>dative</i> |
|-------------|----------|--------------|----------------|---------------|
| | <i>1</i> | ...-li | sa-... | (s)am-... |
| <i>Sing</i> | <i>2</i> | is(h)-... | chi-... | chim-... |
| | <i>3</i> | ∅ | ∅ | im-... |
| | <i>1</i> | (k)il-... | po-... | pom-... |
| <i>Plur</i> | <i>2</i> | has(h)-... | hachchi-... | hachchim-... |
| | <i>3</i> | ∅ | ∅ | im-... |

(Payne 1982:359)

Interestingly, these three affixes never occur all three together, there are maximally two affixes found on a verb (Payne 1982:367).⁸ If there are more than two candidates, a choice has to be made for maximally two of them. There are a few restrictions on this choice. First, two patients, two datives, or a patient and a dative both may not occur on any one verb. Combined with the zeros of the third person (see Table 5) this is in most cases enough to bring the marking down to two affixes, as shown in (10a) and (10b). If there are two affixes present, one of them consequently has to be an agent-affix, the other can be either patient or dative, as shown in (11a) and (11b). Concluding, maximally two affixes are present, but the two affixes do not have a fixed role.

- (10) a. am-ambi-tok
1SgDat-kill-Past
‘He killed it for me’ (Payne 1982:358)
- b. chim-a-li-tok
2SgDat-give-1SgAgt-Past
‘I gave it to you’ (Payne 1982:358)
- (11) a. is-sa-shō’ka
2SgAgt-1SgPat-kiss
‘You kissed me’ (Payne 1982:355)
- b. is-sam-ollaha’li
2SgAgt-1SgDat-smile
‘You’re smiling at me’ (Payne 1982:358)

There are a few languages where it is possible to mark three different roles inflectionally on the verb. But even in the few languages that allow for the possibility, the constructions are rather marked. The first example is from Yimas (Sepik-Ramu, New Guinea). In Yimas, the three core cases are only marked inflectionally in the special combination of third person object, third person non-singular indirect object and first or second person subject (sentences like ‘I give it to them’), an example is shown in (12). In all other constructions with three different core-participants there are maximally two of them overtly marked.

- (12) *uraŋ* *k-mpu-ŋa-tkam-t*
 coconut **Cl.6-3PIAgt-1SgDat-show-Perf**
 ‘They showed me the coconut’ (Foley 1991:213)

Two other languages that are normally analyzed as showing three different sets of inflectional pronominal marking are the Caucasian neighbours Abkhaz (North Caucasian, Georgia) and Georgian (South Caucasian, Georgia). These languages are in geographical contact, but do not have a direct linguistic relationship. In Abkhaz it indeed occurs that three different core cases are marked on a verb, in Georgian, though, it does not occur.

In Abkhaz – a morphologically ergative language – the object (i.e. absolutive) often disappears if an indirect object is present in the meaning, as shown in (13). The prefix is realized here as zero, although there is an overt third person plural absolutive, ‘*rə/də-...*’ (Hewitt 1979:101).

- (13) (*sarà*) *a-x^oəč’-k^oà* *a-š^oq^o’-k^oà* *∅-rə-s-to-yt’*
 (I) Def-child-Pl Def-book-Pl **3PIAbs-3PIIObj-1SgErg-give-Fin**
 ‘I give the books to the children’ (Hewitt 1979:105)

There are also examples in the grammar where all three affixes co-occur with one verb. Two examples are shown in (14). It remains unclear what regulates the distribution of this abundant marking. Note that example (14a) leaves a rather ‘constructed’ impression as people are not normally given to someone else.

- (14) a. *d-bə-l-ta-r+t^o’*
 3SgMascAbs-2SgIObj-3SgFemErg-give-Purp
 ‘She gives him to you’ (Hewitt 1979:171)
- b. *yə-b-sə-r-q’a-c’e-yt’*
 3SgNeutAbs-2SgIObj-1SgErg-Cause-Prev-Do-Fin
 ‘I made you do it’ (Hewitt 1979:171)

The Georgian situation seems to resemble Abkhaz as there are also three different inflectional paradigms, shown in Table 6.⁹ The main difference though is that these three affixes never overtly co-occur together on one verb. Note beforehand that there are a few zeros in the table, which leaves only a few possibilities for three core cases to be marked together. And note that the third person indirect object is often zero, for some speakers even always (Harris 1981:29).

Table 6: Georgian singular pronominal affixes

| | | <i>subject</i> | <i>object</i> | <i>ind. object</i> |
|-------------|----------|----------------|---------------|--------------------|
| | <i>1</i> | v-... | m-... | mi-... |
| <i>Sing</i> | <i>2</i> | ∅-... | g-... | gi-... |
| | <i>3</i> | ...-s | ∅-... | u/∅-... |

(Harris 1981:29; Kathman 1995:154)

There still remain a few combinations that could yield three affixes. One of them is the combination v-g-u (1Subj-2Obj-3IndObj ‘I give you to him’; semantically awkward, but alas). However, there happens to be a regular deletion rule (15) which prevents this combination (Harris 1981:31).

(15) [v-...] → ∅ / __ [g-...]

Only combinations of a third person subject with first or second person object remain as possible candidates for verbs with three affixes. But here a special rule of Georgian rules out these combinations. Georgian has so called ‘object camouflage’: the first or second person direct object is demoted to possessor-phrase if there is an indirect object (Harris 1981:31). This demotion is shown in (16); (16a) is the ungrammatical construction with three core cases marked on the verb, (16b) is the grammatical counterpart with the object participant demoted to a possessor-phrase.

(16) a.* Vano (šen) g-a-dereb-s Givi-s
 Name (you) 2SgObj-3SgIndObj-compare-3SgSubj Name-Dat
 ‘Vano is comparing you to Givi’ (Anderson 1984:208)

b. Vano šens tav-s a-dareb-s Givi-s
 Name your self-Acc 3SgIndObj-compare-3SgSubj Name-Dat
 Vano is comparing you to Givi. (Anderson 1984:208)

Georgian may seem to offer three core cases, but maximally two are marked on a verb. In general, it is rather uncommon to see verbs with more than two participants

marked, uncommon cross-linguistically, as well as uncommon within a particular language. This indicates that a theory of grammatical core case marking should make a distinction between the first two participants to be marked, and the rest. Until the first two participants, it is perfectly normal cross-linguistically for them to be marked inflectionally on a verb. More than two inflectionally marked participants is rather uncommon.

Conclusion

When pronominal marking is pursued from a cross-linguistic point of view, things get more complicated, but some problems also loosen up. The variability of human language turns out to be much greater than one may have expected from the – rather restricted – variation in the structure of European languages. Consequently, there turns out to be much more to explain and to acknowledge before any viable universal claims are possible. On the other hand, there are a few aspects where easy points can be scored. To be able to compare a wide variety of languages, it is necessary to clearly define certain notions that are often taken for granted. The definition for ‘pronominal system’ used in this paper was designed to be applicable to (almost) all known linguistic structures. If this definition is taken serious, then a few ‘universal’ properties become almost inevitable. Languages tend – with overwhelmingly more than chance frequency – not to be redundant in their pronominal marking. Next, languages normally mark pronominal reference maximally once, double marking is highly marked – with overwhelmingly more than chance frequency. Combining these two insights, double marking does not function as noise reduction, but actually adds information when it is needed. Thirdly, it is highly unusual that more than two participants are marked inflectionally on a verb. If there are more participants in the discourse, then some of those will not be marked overtly – again with overwhelmingly more than chance frequency.

References

- Anderson, S. R. (1984) ‘On Representations in Morphology: Case Marking, Agreement and Inversion in Georgian’ *Natural Language and Linguistic Theory* 2(2): 157-218.
- Birk, D. B. W. (1976) *The Malakmalak Language*. Canberra, Australian National University.
- Bowden, J. (1998) ‘Cooking up trouble with Taba: theoretical issues arising from Taba applicative constructions’. paper read at: Argument-Structure Group Meeting, Max Plank Institute Nijmegen: March 31, 1998.
- Broschart, J. (1997) ‘Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs’ *Linguistic Typology* 1(2): 123-166.
- Carlson, R. J. (1990) *A Grammar of Supyire, Kampwo Dialect*, Dissertation, University of Oregon.

- Cooke, J. R. (1968) *Pronominal Reference in Thai, Burmese, and Vietnamese*. Berkeley, University of California Press.
- Cysouw, M. (to appear-a) 'Ik, jij ... en verder? Over de variatie aan persoonsdeixis in de talen van de wereld' *Gramma/TTT*.
- Cysouw, M. (to appear-b) 'Interpreting Typological Clusters' *Linguistic Typology*.
- Dixon, R. M. W. (1994) *Ergativity*. Cambridge, Cambridge University Press.
- Durie, M. (1985) *A Grammar of Acehnese*. Dordrecht, Foris.
- Fahner, C. (1979) *The Morphology of Yali and Dani*, Dissertation, Rijksuniversiteit te Leiden.
- Fernandez Garay, A. (1993) *Le Tehuelche, description d'une langue en train de disparaître*, Dissertation, Sorbonne, Paris.
- Foley, W. A. (1991) *The Yimas Language of New Guinea*. Stanford, Stanford University Press.
- Forchheimer, P. (1953) *The Category of Person in Language*. Berlin, Walter de Gruyter.
- Givón, T. (Ed.) (1983) *Topic Continuity in Discourse: A Quantitative Cross-language Study*. Amsterdam, John Benjamins.
- Harris, A. C. (1981) *Georgian Syntax: A study in Relational Grammar*. Cambridge, Cambridge University Press.
- Hewitt, B. G. (1979) *Abkhaz*. Amsterdam, North Holland Publishing Company.
- Ingram, D. (1978) 'Typology and Universals of Personal Pronouns' in: J. H. Greenber (Ed.) *Universals of Human Language*. Stanford, Stanford University Press. 3: word structure: 213-248.
- Innes, G. (1966) *An Introduction to Grebo*. London, School of Oriental and African Studies.
- Kathman, D. (1995) 'Verb Agreement and Grammatical Agreement' in: C. S. Burgess, K. Dziwirek and D. Gerdts (Eds.) *Grammatical Relations, theoretical approaches to empirical questions*. Stanford, CSLI Publications: 153-170.
- Lehmann, C. (1988) 'On the Function of Agreement' in: M. Barlow and C. A. Ferguson (Eds.) *Agreement in Natural Language*. Stanford, Center for the study of Language and Information: 55-66.
- Mithun, M. (1991) 'Active/agentive case marking and its motivations' *Language* 57: 510-546.
- Moravcsik, E. A. (1978) 'Agreement' in: J. H. Greenberg (Ed.) *Universals of Human Language*. Stanford, Stanford University Press. 4. *Syntax*: 331-374.
- Moravcsik, E. A. (1988) 'Agreement and Markedness' in: M. Barlow and C. A. Ferguson (Eds.) *Agreement in Natural Language*. Stanford, Center for the study of Language and Information: 89-106.
- Mühlhäusler, P. and R. Harré (1990) *Pronouns and People: the Linguistic construction of social and personal identity*. Oxford, Basil Blackwell.
- Munro, P. and L. Gordon (1982) 'Syntactic Relations in Western Muskogean' *Language* 58: 81-115.
- Pawley, A. (1993) 'A language which defies description by ordinary means' in: W. A. Foley (Ed.) *The Role of Theory in Language Description*. Berlin, Mouton de Gruyter: 87-130.
- Payne, D. L. (1982) 'Chickasaw Agreement Morphology, a Functional Explanation' in: P. J. Hopper and S. A. Thompson (Eds.) *Studies in Transitivity*. New York, Academic Press: 351-378.
- Rose, S. M. and B. F. Carlson (1984) 'The Nootka-Nitinaht Passive' *Anthropological Linguistics* 26(1): 1-12.
- Rottland, F. (1982) *Die Südnilotischen Sprachen*. Berlin, Dietrich Reimer.

- Tucker, A. N. and M. A. Bryan (1966) *Linguistic Analyses of the Non-Bantu Languages of North-Eastern Africa*. London, Oxford University Press.
- Wells, M. A. (1979) *Siroi Grammar*. Canberra, Australian National University.
- Wierzbicka, A. (1996) *Semantics: Primes and Universals*. Oxford, Oxford University Press.

Notes

¹ Non-occurrence of types indeed only *indicates* restrictions on possible structures of human language. It is still a matter of debate which explanation to choose for skewing in frequencies. Besides universal properties of human language, the skewing could also be a result of large-areal bias, cf. Cysouw (to appear-b).

² All lesser known languages that appear as examples in this paper are, on first appearance, followed by some classifying information between brackets: first, the linguistic family this language belongs to, and, second, the country where this language is mainly spoken. I do not have any political intentions with this information and I regret any harm I inflict on the reader by using an ‘incorrect’ classification or nationality. If I did, it is because of ignorance, not of obstinacy.

³ The paradigmatical variation is dealt with elsewhere, cf. Cysouw (to appear-a).

⁴ One fixed aspect is that the pronominal elements always relate to a *sentence-predicate*. This is not because of any universal characteristic of human language, but because I have defined the domain as such. As described in the methodology section, I disregard pronominal possession and other adnominal marking here.

⁵ This is different from arguing whether there is a difference between nouns and verbs in Nootka – a much debated question. If the marking of person would be taken as the decisive argument in this debate, then there would indeed not be a difference between nouns and verbs. The defining difference between nouns and verbs could also be another characteristic, cf. Broschart (1997).

⁶ English of course has double marking of third person singular. For reasons of cross-linguistic comparability I have chosen to define a pronominal system on the basis of first and second person marking, and, consequently, English does not have an inflectional pronominal system by this definition.

⁷ If agreement alternatively is defined basically as a relation between a full noun and a pronominal element, then of course agreement occurs regularly in the languages of the world. Even in these cases though there is a case to be made that sentences with a full noun are in a sense ‘marked’ as most languages tend to leave out any nominal reference as soon as the communicational setting does not need it – in case all addressees know what the speaker is talking about.

⁸ Note that the dative paradigm is probably historically derived from the patient paradigm by adding ‘-(i)m-...’. Synchronically though they are two different paradigms.

⁹ Only the singular forms are shown in the table, as there are some complicating factors in the plural which are not of interest for the present discussion.