The Phonological Composition of Personal Pronouns: Implications for Genetic Hypotheses

Author(s): Matthew J. Gordon

Proceedings of the Twenty-First Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Historical Issues in Sociolinguistics/Social Issues in Historical Linguistics (1995), pp. 117-128

The Phonological Composition of Personal Pronouns: Implications for Genetic Hypotheses

Matthew J. Gordon
University of Michigan

1. INTRODUCTION. Personal pronouns[1] can be found in virtually every language in the world. In many languages they are used with remarkable frequency and serve a variety of functions. Although linguists from almost every subfield have discussed the unique role that pronouns play, very little work has been done to determine what a typical pronoun or pronoun system looks like. This is particularly true with regard to their phonological make-up. While it may not be immediately apparent why we should be concerned with this gap in the research on pronouns, this issue has recently attracted the attention of historical linguists, in particular those involved in attempts to establish or refute claims of language relatedness.

Personal pronouns are, in many languages at least, more stable than other elements. They are reported to be more resistant to semantic change and less likely to be replaced through borrowing.[2] Because of their putative stability, they, along with other 'core' vocabulary items, often figure as evidence in proposals for the genetic affiliation of languages. Recently, however, some questions have been raised about the use of pronoun data in such proposals, particularly when these proposals are arrived at through the method of 'mass lexical comparison' associated with Joseph Greenberg. This method draws its genetic conclusions primarily on the basis of shared similarities of form and meaning across languages. Unlike traditional approaches, Greenberg's method does not rely on systematic sound correspondences or comparative reconstruction. Because of this, many historical linguists believe mass comparison to be incapable of distinguishing similarities due to genuine inheritance from those due to chance or other non-genetic explanations. Opponents of Greenberg's classification of American Indian languages (1987) have been especially critical on this methodological point (see Campbell 1993, Thomason 1990). Although almost every aspect of Greenberg's 'Amerind' proposal has been questioned, some of the strongest opposition has been focussed on the pronominal evidence. Criticism of these data relies in part on some generally accepted but rarely tested notions regarding the phonological composition of pronouns. Specifically, it has been claimed that (1) pronouns (and other grammatical morphemes) tend to involve a limited number of segments from a language's phonemic inventory, (2) the same sounds reappear in pronouns from apparently unrelated languages, and (3) there is a non-arbitrary relationship between certain sounds and certain pronominal meanings. These claims seem to suggest that special factors operate in determining the phonological shape of pronouns and that one must be cautious, therefore, in offering pronoun data to support hypotheses of language relatedness.

This paper uses data collected from a cross-linguistic survey of pronouns to explore these claims. Following a brief methodological overview, each of the three claims is examined and tested. In discussing the results of these tests, the focus is on their implications for the problem of establishing genetic relationships.

2. METHODOLOGY. The present research was conducted as part of a broad study which investigated phonological as well as morphological, semantic and syntactic aspects of pronouns and pronoun systems. The goal of this project was not to establish a typology of pronouns (i.e. to define the limits of variability by outlining what is and is not possible), but rather to provide some guidelines as to what is typical of pronouns. This objective necessitated that many languages be considered and that the data from these languages be as independent as possible. Therefore, my sample included 62 languages selected to be both genetically and areally diverse.[3] (see Appendix for a list of languages with their genetic affiliations and locations).

The pronominal data collected included both independent and bound morphemes. A form was taken to be pronominal if it functioned to mark the grammatical category of person.[4] In addition to the actual pronominal forms, an inventory of phonemes was taken for each language in the sample. These inventories were used not only to compare the range of sounds available to each language but also to establish a uniform transcription system.

3. RESULTS AND DISCUSSION

3.1. LIMITED USE OF PHONEMIC INVENTORY. Pronouns are grammatical ('function'), as opposed to lexical ('content'), morphemes. They serve the grammatical function of marking *inter alia* person. Certain phonological properties are said to be characteristic of grammatical morphemes. We are most concerned here with the claim made by several authors that languages tend to use only a limited number of sounds from their total segment inventory in the formation of grammatical morphemes. For example, Floyd (1981) finds only seven (m n t t^h k r s) of Classical Greek's fifteen consonants occurring in inflectional forms and mentions similar constraints for Hebrew, German, Latin and English.[5] If such restrictiveness is, in fact, a widespread phenomenon, historical linguists must use greater caution in judging similarities in grammatical forms because, with fewer sounds to choose from, there is a higher probability of random (i.e. non-genetic) resemblances across languages.[6]

I tested this hypothesis as it relates to pronouns by comparing the phonemic inventory of each language in the sample to the sounds appearing in the database of pronominal forms from that language; that is, for each language I counted the number of phonemes available in inventory and the number that appeared in the language's pronouns.[7] A summary of the results of these comparisons is found in Table 1, which shows the range of variation and averages across the 62 language sample for inventory size, the number of phonemes used in pronouns, and the ratio of the number of phonemes used to the number in inventory.

	Inventory		Used		% Used	
	Range	Avg.	Range	Avg.	Range	Avg.
Consonants:	7-88	22.2	5-15	9.4	10.7-100%	49.5%
Vowels:	3-17	7.1	3-10	5.1	23.5-100%	78.4%
All Sounds:	10-94	29.3	8-22	14.5	17.2-100%	54.4%

Table 1: Summary comparison of phonemic inventory size and number of phonemes appearing in pronominal forms from a sample of 62 languages. Figures indicate number of segments or percentages of phonemes used as a ratio of total inventory.

As Table 1 shows, languages vary greatly with regard to their phonemic inventories. The largest inventory was found in Tsaxur (NE Caucasian) with its 88 consonants and 6 vowels and the smallest in Piraha (Paezan) which has only 7 consonants and 3 vowels.[8] Actually, these extremes suggest greater variation than there really is, as the standard deviation was 11.9 segments and over 77% of the languages have inventories in the range of 20-37 segments. The overall average inventory size was 29, which is in line with the average of 31 obtained by Maddieson (1984) with a much larger sample.[9] Turning to the data on sounds used in pronoun forms, we see much less variation, with an average of 14.5 segments and a standard deviation of 3.3. Khmer made use of the greatest number of sounds (22), while both Mandarin and Piraha used only 8. In some instances there seems to be a direct connection between the number of sounds used and the size of the language's inventory; however, there are many exceptions. A linear regression analysis found inventory size to be an unreliable predictor of the number of sounds used ($r=.193$, $f(1,60)=2.312$, $p=.134$). It seems that many languages use approximately the same number of sounds in pronouns despite variation in their inventory size. This is confirmed, in part, by the wide range of the ratios in Table 1. Languages with large inventories tend to employ only a small percentage of their sounds (e.g. White Hmong 11/64 = 17.2%), while those with small inventories employ a much greater percentage (e.g. Seneca 16/16 = 100%). In either case, most languages (over 75% in this sample) use between 10 and 17 sounds in pronoun forms.

These results are intriguing and indicate that there may in fact be some type of limitation on the number of sounds appearing in pronouns. However, the consistency in these numbers suggests a more pragmatic explanation than the mysterious restriction implicit in the claim that languages make only limited use of their phonemic inventories for grammatical forms. First of all, pronouns, and grammatical morphemes in general, tend to be shorter (i.e. contain fewer segments) than other elements in a language. As a practical explanation for this phenomenon, one might suggest that such morphemes need to be short because they appear so frequently. Whatever the explanation, this tendency has the net effect of limiting the range of sounds used in pronouns by simply reducing the number of available slots. Similarly, for any given language, the number of sounds used is clearly influenced by the number of forms collected. In the present sample there was tremendous variation in the number of pronominal forms available in the languages, from as few as 8 to as many as 232 forms. Obviously, having fewer forms reduces a language's capacity to display its sounds.[10] Not surprisingly, then, the correlation between the number of forms and the number of sounds used proved to be quite strong when calculated through linear regression ($r=.43$, $f(1,60)=13.58$, $p<.0005$).

It must be pointed out that these results do not necessarily disprove the notion that some special property of pronouns restricts the number of sounds used in their composition. I have merely suggested that the length and number of pronominal forms may provide another, less interesting, explanation. This explanation could be tested more thoroughly by comparing the phonemes used in pronouns with those used in non-grammatical morphemes of the same length and frequency. If languages can be shown to use a comparable repertoire of sounds in both contexts, it would be more difficult to maintain that pronouns are uniquely selective.

In sum, we have found that many languages do employ only a partial set of their phonemic inventories in pronoun forms. The original claim does not, however, hold universally, since there are languages (2 in this sample) that use all of the sounds available to them and others that use nearly all their sounds (e.g. Tiwi 17/18; N. Sierra Miwok 18/20). More importantly, however, even in those cases where a limited number of sounds appear in pronominal forms this fact appears to be a predictable consequence of relatively straightforward aspects of the data, such as the number of sounds in inventory, the length of the forms, or the number of forms available.

3.2. THE PHONEMES USED IN PRONOUN FORMS. In addition to pronouns' using only a limited number of sounds, it is claimed that the same types of sounds tend to reappear in pronouns of unrelated languages (see e.g. Campbell 1993). This tendency, if true, is important for historical linguists to note because it provides a non-genetic explanation of phonetic similarities among comparable morphemes of different languages. The validity of this claim was tested with this set of 62 languages by simply counting the number of languages using each of the various sounds found in the database. In the present sample 95 different sounds appeared in pronominals: 26 vowels and 69 consonants.[11] As expected, most (54%) of these sounds appeared in only one or two languages,[12] while a smaller set of sounds was found to recur in many languages. The most frequently used sounds, those appearing in at least 10% of the languages, are listed in Table 2.

Consonants

	n	%		n	%
n	58	93.5	p	22	35.5
m	47	75.8	l	20	32.3
k	44	71.0	ʔ	19	30.6
t	42	67.7	d	17	27.4
j	33	53.2	tʃ	15	24.2
w	27	43.5	b	14	22.6
h	25	40.3	ɲ	12	19.4
ŋ	24	38.7	ʃ	12	19.4
s	23	37.1	g	11	17.7
ɾ	23	37.1			

Vowels

	n	%		n	%
a	61	98.4	ɛ	17	27.4
i	56	90.3	ə	14	22.6
u	43	69.4	ɪ	9	14.5
o	35	56.5	ɔ	9	14.5
e	32	51.6	ɨ	8	12.9

Table 2: The most frequently used sounds in pronominal forms. Indicated are the number of languages (n) using each sound and the percentages they represent of the total sample of 62 languages

The results in Table 2 support the claim that certain sounds can be expected to appear in pronoun forms, but again we must be careful in assessing the significance of this fact. When we examine the data, we find that there is nothing very

surprising about the sounds listed in Table 2. The vowels are all quite frequent in the languages of the world: 7 of the 10 (a, i, u, o, e, ɛ, ɔ) constitute Maddieson's list of most common vowel qualities (1984:125). Similarly, the consonants are also among the most commonly occurring in the world. In fact, they are all included by Maddieson in his 'modal' inventory, a set comprising the 20 most frequent consonants found in his 317 language sample (Maddieson 1984:12).[13] These facts beg the question of whether the sounds found in Table 2 are the most frequent in the data simply because they are the most frequent cross-linguistically or whether their frequency is somehow relatable to their appearance in pronouns.

Although the simplest solution may be to explain the data in Table 2 as a reflection of general markedness patterns (as measured by cross-linguistic frequency), other factors may have some explanatory value and should not be overlooked. One such factor that deserves brief mention is the notion of perceptual salience. The basic idea is that since grammatical morphemes typically contain only a few segments, a language must get as much out of each segment as possible. One way of achieving this goal is to choose the most perceptually salient sounds in these contexts. Campbell (1993:4) offers this as an explanation for the common appearance of nasals in grammatical markers and cites Maddieson's (1984:70) observation that nasals 'are rarely subject to confusion with other types of consonants'. There is a problem with this argument, however, which Campbell does not consider. While nasals as a class are very easy to distinguish from other sounds, and hence more salient, there is often confusion within the class differentiating one nasal from another (see studies cited in Maddieson 1984:70). If salience were the key factor, then we would expect languages to use only one nasal, but in fact this does not seem to be the case. Of the 60 languages in my sample that used nasals only 5 (8.3%) chose to employ a single nasal. The limitations of the salience argument are evident in the following pronominal paradigm from Katla, a Kordofanian language:

	Sing.	Plural
1st per.	ɲɔŋ	nɛn
2nd per.	ŋaŋ	nɔn
3rd per.	ŋuŋ	ŋiŋ

While these data may represent an extreme in terms of both nasal usage and overall perceptual similarity, it is not uncommon to find only minor phonetic differences distinguishing two or more forms within the same paradigm. This suggests that the need for grammatical forms to be clearly distinguished may not be as great as has been assumed. Perceptual salience may play some role in determining the phonological composition of pronouns, but there seems to be no reason to believe that its influence is greater in pronoun systems than in other areas of the language.

Despite this negative assessment, we should not rule out completely the possibility that certain other phonetic and phonological properties make some sounds particularly well suited to function in pronouns and other grammatical morphemes. Various types of explanations have been suggested, such as the claim that dentals and sonorants are preferred in bound morphology because they readily participate in clusters (see Callaghan 1991:53). These claims cannot be explored in this paper, but certainly merit further consideration.

3.3. CONNECTIONS OF SOUND AND MEANING. The final issue to be addressed involves the possibility of a non-arbitrary relationship between the meanings denoted by pronouns and the sounds used to compose them. Evidence for such a relationship would be of interest to historical linguists as another non-genetic way of accounting for similarities across languages. Thomason (1990:9) raises this issue in her criticism of Greenberg's use of pronoun data. Among her objections she lists the widespread 'affective and onomatopoetic uses of nasals' and mentions the examples of *mama* and *nana* as kin terms, but she gives no indication of how pronouns are affected by such sound symbolic patterns. Another of Greenberg's critics, Campbell, is much more explicit in his formulation of this claim. Borrowing an idea originally presented by Goddard (1986), Campbell suggests that the pattern of *n* marking first person (which is posited for Greenberg's Amerind family) may have a basis in child language. Since Campbell claims this argument has been misrepresented by Greenberg and his supporters, I quote it in full:[14]

> [C]hild-language expressions around the world abound in self-directed and other-directed words containing nasal consonants. The ultimate reason for this is the universal physical fact that a gesture equivalent to that used to articulate the sound *n* is the single most important voluntary muscular activity of the nursing infant. As Goddard (1986:202) points out, possibly this factor and the tendency for primary grammatical morphemes to consist of a single, unmarked (phonetically commonplace) segment account for the widespread appearance of *n-* in 'first-person' pronouns. Incidentally, in many societies, particularly among hunting and gathering groups, infants may continue to nurse until the age of five, sometimes longer, well into and beyond the age of language-acquisition. (Campbell 1993:6)

This statement suggests that Greenberg's Amerind evidence is merely a reflection of a universal preference for first person markers using *n*. If this preference really exists, we should expect to find *n* unusually frequent in the first person as opposed to the other two persons.

This claim can be investigated in the present sample by counting the number of languages employing *n* in each of the three persons. This approach presents a complication, however, due to the broad nature of the database. Since the database contains all the pronominal forms from the sample languages, counting all the data increases the amount of phonological material having nothing to do with person marking. For example, suppose a language has a different set of pronouns for each of several grammatical cases, and suppose that one of these cases is formed by adding the suffix *-n* to the pronominal stems for each person/number combination. If all the data are considered, that *n* would be counted for each person. In some instances this would simply add unnecessarily to the number of languages with a certain sound in each of the three persons; however, in many instances the net effect would be to obscure any real correlations that might otherwise be evident. Returning to our example, suppose the first person was the only one of the three in which an *n* appeared in the stem. This potentially interesting fact would be lost because *n* would be recorded for each person, due to its appearance as a case marker. For this reason a single set of pronouns was selected for analysis from each language. The goal in choosing these sets was to eliminate as much unnecessary phonemic overlap as possible, while preserving cross linguistic comparability. For most languages this meant choosing the least marked set of independent pronouns (e.g. the nominative or absolutive set); however, in a few instances bound person markers were selected, particularly when they were obviously segmentable from a common pronominal stem. Using these 'basic' sets

of pronouns, I measured the frequency of occurrence of all sounds across the three grammatical persons, with the idea that even if I found no significant pattern for *n*, I might for other sounds. The results for the most frequent sounds (those appearing in over 50% of the languages (see Table 2)) were distributed as follows:

	Person		
	1st	2nd	3rd
a	42	37	42
i	38	34	31
u	16	16	22
o	11	9	17
e	14	15	22
n	34	30	32
m	25	16	11
k	19	18	17
t	10	14	24
j	11	8	8

Table 3: Frequency of sounds occurring in each grammatical person for the basic sets of pronouns. Figures represent number of languages out of 62 total.

It is quite evident from the consistency of these figures that most sounds seem to be distributed independently of the category of person. The data in Table 3 were subjected to chi-square testing and with two exceptions none of the patterns was found to be significant beyond the level of .05. The exceptional cases are *m*, which shows a preference for first person (chi-square = 8.06, df = 2, p = .018), and *t*, which correlates with third person (chi-square = 8.76, df = 2, p = .013).

The implications of these results for the specific claim made by Campbell and Goddard are quite apparent: there is simply no indication of an overall preference for *n* in first person markers. More importantly, my data give some indications that the distribution of *n* within the members of the proposed Amerind group is somewhat unusual. This is clear when the data for *n* are separated to compare the 17 Amerind languages with the others:

	1st	2nd	3rd	X^2	p
Amerind	10	4	8	4.47	.107
Others	24	26	24	.239	.887

Table 4: Frequency of *n* occurring in each grammatical person for the basic sets of pronouns from 17 Amerind and 45 non-Amerind languages

Although neither of these patterns is statistically significant, it is obvious that the data are much more evenly distributed outside of the Amerind group. This observation should not, however, be taken as strong support for the validity of Amerind as a genetic group. We are dealing with the results of a superficicial phonological analysis of a limited amount of data from a small set of languages. Even if we are willing to overlook these limitations, we still have no reason to claim that the patterns demonstrated are due to common genetic inheritance. Nevertheless, if further studies should confirm a significant connection of first person forms with *n* in the Amerind languages, then this fact would need to be

explained and it is clear that the situation cannot be attributed to some universal preference.

The fact that a significant correlation was found linking *m* and 'first person' raises some interesting questions. Perhaps Campbell and Goddard were essentially correct in their claim but simply chose the wrong nasal. After all, it seems plausible that *m* could be incorporated into their child language explanation.[15] There is a complication, however, involving another of Greenberg's genetic proposals, namely his Eurasiatic group. For this group Greenberg has claimed that *m* is the characteristic marker of first person (see Greenberg 1991). Given the results discussed here we may wonder whether Greenberg has simply observed a general linguistic tendency that has nothing to do with genetic inheritance. To test this, we may simply compare the distribution of *m* across the three persons with and without the Eurasiatic languages in the sample. The original sample contains 9 members of the Eurasiatic group, five of which have *m* in the first person. With so few languages we cannot reliably determine the significance of the distribution within the proposed family, but a chi-square test was performed for the other languages:

	1st	2nd	3rd	X^2	p
Eurasiatic	5	1	1	--	--
Others	20	15	10	4.65	.098

Table 5: Frequency of *m* occurring in each grammatical person for the basic sets of pronouns from 9 Eurasiatic and 53 non-Eurasiatic languages

The original distribution of *m* was significantly weighted in favor of the first person, but when the Eurasiatic languages are not included, this correlation is no longer statistically significant. Still, the general pattern remains: *m* is more frequent in first person than in the other two. If more languages were sampled it seems likely that a significant pattern could be confirmed.[16] This suggests that there is indeed some general preference for *m* to appear in first person forms. Whether this preference is ultimately connected to child language phenomena or to some other explanation remains an open question.[17] Certainly, we will need to investigate new avenues to explain the other significant pattern, the one linking *t* and third person, although, at this point, I know of none that has been offered.

4. CONCLUSIONS. Various claims have been made to suggest that pronouns merit special consideration by linguists attempting to establish genetic relations among languages. Their phonological composition is said to derive from a restricted subset of the language's inventory of sounds, a small set of unmarked sounds is thought to be common to pronouns of diverse languages, and certain pronominal meanings are claimed to have a predilection for certain sounds. Each of these assertions has been investigated in this paper using data collected from a controlled sample of languages. Although the results have been mixed, they offer clear methodological implications for genetic linguistics.

The data considered here confirm that there is a tendency, though certainly not universal, for languages to employ a limited number of their phonemes in pronoun forms. This really should come as no surprise, however, since any subset of data from a language is likely to display only a partial amount of the complete phonemic inventory. Similarly, we are not surprised to find that the sounds which recur in the pronouns of language after language are also among the least marked, most

cross-linguistically frequent sounds in general. While these observations suggest that there are no mysterious factors operating to determine the phonological composition of pronouns, the uncovering of two significant patterns correlating sound and meaning (*m* and *t* with first and third person respectively) raises new questions. Further research is needed to confirm and seek explanations for these patterns, but at this point it is valuable to be aware of such connections and consider the possibility that they may result in some non-etymological similarities across languages.

Actually, the same can be said of the other findings as well. If, for whatever reason, the number of sounds used in pronouns is limited and certain sounds are generally more frequent, we can certainly expect a greater incidence of chance matchings that do not reflect any common ancestry. The methodological lesson to be drawn here is that the burden of proof in such cases is necessarily greater (cf. Ringe 1992). When considering cross-linguistic resemblances of form and meaning, we can never prove that the genetic hypothesis is the only explanation, the best we can hope for is to establish that it is the most reasonable explanation. In working with pronouns, it seems this job is all the more difficult.

APPENDIX: Languages used in this study

Language (Location)	Genetic Affiliation*
Alamblak (Papua-New Guinea)	Sepik-Ramu; Indo-Pacific
Amele (Papua-New Guinea)	Madang; Trans New Guinea; Indo-Pacific
Apalai (N Brazil)	Carib; Amerind
Arabic (Persian Gulf)	Semitic; Afro-Asiatic
Barasano (Colombia)	Tucanoan; Amerind
Bashkir (Bashkir Rep.)	Turkic; Altaic; Eurasiatic
Basque (N Spain)	Isolate (possibly Dene-Caucasian)
Bobo (Burkina Faso)	Mande; Niger-Kordofanian
Bukiyip (N Papua-New Guinea)	Torricelli; Indo-Pacific
Cahuilla (S California)	Uto-Aztecan; Amerind
Canela-Kraho (Central Brazil)	Je; Amerind
Delaware (NE Coast America)	Algonquian; Almosan; Amerind
English (USA)	Germanic; Indo-European; Eurasiatic
Ewe (Togo)	South Central Niger-Congo; Niger-Kordofanian
Fula (Senegal)	W.Atlantic; Niger-Congo; Niger-Kordofanian
Georgian (Georgia)	Kartvelian
Greenlandic (W Greenland)	Eskimo-Aleut; Eurasiatic
Gujarati (W India)	Indic; Indo-European; Eurasiatic
Hmong (White) (N Thailand)	Miao-Yao; Austric
Hua (Papua-New Guinea)	East New Guinea Highlands; Indo-Pacific
Hualapai (Arizona)	Yuman; Amerind
Ika (Colombia)	Chibchan; Amerind
Itelmeny (Kamchatka Penn.)	Chukchi-Kamchatkan; Eurasiatic
Jacaltec (Guatamala)	Mayan; Amerind
Kanuri (Nigeria)	Western Saharan; Nilo-Saharan
Katla (S Sudan)	Kordofanian; Niger-Kordofanian
Khmer (Cambodia)	Mon-Khmer; Austroasiatic; Austric
Kikuyu (Kenya)	Central Bantu; Niger-Kordofanian
Koasati (Lousiana)	Muskogean; Amerind
Korean (Korea)	Altaic (?); Eurasiatic
Kunama (Eritrea)	Nilo-Saharan
Lahu (China/Myanmar)	Lolo-Burmese; Tibeto-Burman; Sino-Tibetan

Maasai (Kenya)	Nilotic; Nilo-Saharan
Mandarin (China)	Sinitic; Sino-Tibetan
Margi (Nigeria)	Chadic; Afro-Asiatic
Mari (Mari (S. Russia))	Finnic; Finno-Ugric; Uralic; Eurasiatic
Mba (Zaire)	North Central Niger-Congo; Niger-Kordofanian
Nama (SW Africa)	Khoisan
Nasioi (Bougainville)	East Papuan; Indo-Pacific
Ngandi (N Australia)	Gunwinyguan; Australian
Ngarinjin (NW Australia)	Wororan; Australian
Olcha (E Siberia)	Tungusic; Altaic; Eurasiatic
Ometo (W Ethiopia)	Omotic; Afro-Asiatic
Paiwan (Taiwan)	Paiwanic; Austronesian;Austric
Paumari (Brazil)	Arawakan; Amerind
Piraha (Brazil)	Paezan; Amerind
Pomo (SE) (N California)	Northern Hokan; Amerind
Quechua (Huallaga) (Peru)	Andean; Amerind
Rawa (Papua-New Guinea)	Finisterre-Huon; Trans-NG; Indo-Pacific
Samoan (Samoa)	E.Malayo-Polynesian; Austronesian; Austric
Sarcee (Alberta)	Athabaskan; Na-Dene
Seneca (New York State)	Iroquoian; Amerind
SierraMiwok (N) (N California)	Penutian; Amerind
Squamish (NW Coast America)	Salishan; Almosan; Amerind
Tamil (S India)	Dravidian
Tiwi (N Australia)	Australian
Tsaxur (N Azerbaijan)	Lezgian; NE.Caucasian; North Caucasian
Urubu-Kaapor (NE Brazil)	Tupi; Amerind
Watjarri (W Australia)	Pama-Nyungan; Australian
Yay (N Vietnam)	Tai; Daic; Austric
Yukaghir (NE Siberia)	Uralic; Eurasiatic
Zulu (South Africa)	Central Bantu; Niger-Kordofanian

* For consistency, information on genetic groupings was taken from Ruhlen (1991), although many groups listed therein are controversial.

NOTES

[1] The term "pronoun" as used throughout this paper refers only to personal pronouns, not to other types (e.g. demonstrative, interrogative), and actually denotes various types of person markers as is explained below.

[2] In Dolgopolsky's list of the fifteen semantic values "most impervious to change" the first and second person markers are ranked first and third respectively (Dolgopolsky 1964).

[3] Although care was taken to ensure genetic diversity, it would be impossible to select 62 languages that all linguists would agree are not demonstrably related. Besides including several members of such long-range proposals as Nostratic and Amerind, in a few instances the sample does contain two members of well-established language families; e.g. Indo-European, Bantu, Austronesian. However, even in these cases, an effort was made to select languages as diverse as possible within the family; hence, for example, English and Gujarati were used from Indo-European.

[4] This broad definition presents problems of comparability in some languages, particularly in East and Southeast Asia, where "pronouns" are clearly derived from nouns. In such cases I had to rely on the author's judgement as to which forms function comparably to pronouns in Western languages.

[5] Although Floyd clearly refers only to inflectional elements, Goddard and Campbell (1994) suggest this phenomenon obtains more generally in "primary grammatical morphemes". As this database contains both bound and free forms, my results will apply to this broader formulation.

[6] This assumes that languages are drawing from roughly the same set of sounds to begin with, a claim taken up below.

[7] Although our focus here is the number of sounds appearing in pronoun forms, actually, for many languages these pronominal data contain a variety of grammatical morphemes (e.g. case endings, tense markers etc.).

[8] Everett (1986) claims, apparently correctly, that Piraha has the smallest recorded phonemic inventory in the world.

[9] Maddieson prefers to talk about a range of segments, 20-37, as being typical, because his mean value of 31 is somewhat positively skewed (1984:7).

[10] The relationship between the number of pronominal forms in a language and the number of sounds appearing in those forms is complicated by the fact that in some cases the forms are clearly related across or even within different paradigms (e.g. *me/my/mine*) and in others their is no relationship. In the former case we would expect much greater overlap in terms of the sounds used.

[11] In a limited number of cases phonetically distinct sounds were counted together for the purposes of comparison. This was only done for three classes of sounds and only with languages in which the particular phonetic difference was noncontrastive. Thus, a language's voiceless aspirated stops were included in a count of voiceless unaspirated stops, if the language did not have two separate voiceless series. Similarly, dental sounds were counted together with alveolars, unless a language distinguished these two places of articulation. Finally, various "r-sounds" (including /r/, /ɹ/ and /ɾ/) were counted with alveolar flaps, unless a language had more than one.

[12] 35 appeared only once, 17 appeared twice.

[13] The only difference between this list and Maddieson's is that the latter includes /f/, which was quite infrequent in my sample, appearing in the pronouns of only 2 languages (3.2%).

[14] In fairness to Campbell it should be pointed out that this explanation appears at the end of a long discussion of a variety of objections and he himself describes this claim as "a minor case, certainly not one I favor most" (1993:footnote 11).

[15] Since their argument was raised in the context of the Amerind discussion, this may have led Campbell and Goddard to focus only on *n*.

[16] As few as two more languages using *m* in the first person would produce a significant result beyond the level of .05.

[17] In his comments following the oral presentation of this paper, John Ohala suggested that the suckling reflex of infants most resembles a velar articulation and, therefore, that if any nasal is to be especially frequent in pronouns due to the influence of child language, we should expect that nasal to be an [ŋ].

REFERENCES

[Due to the limited space available, the data sources for the 62 languages on which this study is based are not listed here.]

Callaghan, Catherine A. 1991. Climbing a low mountain. A Festschrift for William F. Shipley, ed. by Sandra Chung and Jorge Hankamer, 47-59. Santa Cruz: Syntax Research Center, Univ. of Calif., Santa Cruz.

Campbell, Lyle. 1993. Problems with pronouns in proposals of remote relationships among Native American languages. Paper read at SSILA meeting, Columbus, Ohio, July 1993.

Dolgopolsky, Aaron B. 1964. Gipoteza drevnejshego rodstva jazykovyx semej Severnoj Eurazii s verojatnostnoj tochki zrenija. Voprosy jazykoznanija 2.53-63. Translated in Typology, Relationship and Time, ed. by V.V. Shevoroshkin and T.L. Markey (1986), 27-50. Ann Arbor: Karoma.

Everett, Daniel. 1986. Piraha. Handbook of Amazonian Languages Vol. I, ed. by D. C. Derbyshire and G. K. Pullum, 200-325. Berlin: Mouton de Gruyter.

Floyd, Edwin D. 1981. Levels of phonological restriction in Greek affixes. Bono homini donum: Essays in Historical Linguistics in Memory of J. Alexander Kerns, ed. by Yoël L. Arbeitman and Allan R. Bomhard, 87-106. Amsterdam: John Benjamins.

Goddard, Ives. 1986. Sapir's comparative method. New Perspectives in Language, Culture, and Personality: Proceedings of the Edward Sapir Centenary Conference (Ottawa, 1-3 Oct. 1984), ed. by William Cowan, Michael K. Foster and Konrad Koerner, 191-214. Amsterdam: John Benjamins.

Goddard, Ives and Lyle Campbell. 1994. The history and classification of American Indian languages: What are the implications for the peopling of the Americas?. Method and Theory for Investigating the Peopling of the Americas, ed. by Robson Bonnichsen and D. Gentry Steele, 189-208. Corvallis, Ore: Center for Study of the First Americans.

Greenberg, Joseph H. 1987. Language in the Americas. Stanford: Stanford Univ. Press.

——. 1991. Some problems of Indo-European in historical perspective. Sprung from Some Common Source: Investigations into the Prehistory of Languages, ed. by Sydney M. Lamb and E. Douglas Mitchell, 125-40. Stanford: Stanford Univ. Press.

Maddieson, Ian. 1984. Patterns of Sound. Cambridge: Cambridge Univ. Press.

Ringe, Jr., Donald A. 1992. On Calculating the Factor of Chance in Language Comparison. Transactions of the American Philosophical Society 82.1. Philadelphia: The American Philosophical Society.

Ruhlen, Merrit. 1991. A Guide to the World's Languages, Volume 1: Classification. Stanford: Stanford Univ. Press.

Thomason, Sarah Grey. 1990. Hypothesis Generation vs. Hypothesis Testing: A Comparison between Greenberg's Classifications in Africa and in the Americas. Language and Prehistory, ed. by Allan Taylor (to appear). Stanford: Stanford Univ. Press.