

# First-person $n$ and second-person $m$ in Native America: a fresh look

Raoul Zamponi

Macerata, Italy <zamponi\_raoul@libero.it>

The presence of a pronominal set with  $n$  in the first person and  $m$  in the second person in numerous Native American languages has been known for more than one century. The number (also approximate) of language families involved and the precise geographical distribution of the  $n$ - and  $m$ -forms, however, are still unclear. In addition, since the  $n : m$  mini-paradigm seems to refer to a historical relatedness which lies beyond the reach of the scientific methods of historical-comparative linguistics, some Americanist linguists currently believe that it is impossible, if not useless, to try to provide a precise explanation of its wide geographical diffusion in the New World. The goals of this article are (i) to examine the presence/absence in all proven linguistic families of the Americas of  $n$ - and  $m$ -pronominal forms for first person or first person singular and second person or second person singular that do not result, as far as it is known, from secondary phonological developments occurred in individual branches; (ii) to reconstruct, as far as it is possible, the geographical distribution of the two forms in pre-European conquest times; and (iii) to attempt to offer a new interpretation of their (apparently robust) geographical diffusion. This article has two appendices published online only: Appendix A (Tables 1-8) and Appendix B (Linguistic data).\*

## 1. Introduction

The presence of a widespread pronominal pattern  $ni$  'I' :  $m$  'you' (sg.) in the languages of Native America was first noticed by the Italian linguist Alfredo Trombetti (Babaev 2009: 37, Manzelli 2015: 302).<sup>1</sup> In his 231-page volume *L'unità d'origine del linguaggio*, published in 1905, Trombetti dedicated four pages (205-208) to the pronouns 'I' and 'you' (sg.) in the American languages then better known with regret that "[t]he list resulted anything but complete for the insufficiency of material at our disposal, but it is certainly enough to give an idea of the great diffusion of these very old and essential elements"<sup>2</sup> (Trombetti 1905: 208). What those four pages

\* I wish to thank Lyle Campbell for commenting on an earlier draft. I am also grateful to two anonymous referees for their assistance in improving this paper and to Harald Hammarström for supplying me with much essential material. The two appendices can be found at [http://www.italian-journal-linguistics.com/wp-content/uploads/Volume\\_29\\_Issue2\\_Appendice\\_Zamponi.pdf](http://www.italian-journal-linguistics.com/wp-content/uploads/Volume_29_Issue2_Appendice_Zamponi.pdf)

present is today known as “multilateral” or “mass” comparison. Pronominal *ni*- and *m*-forms for ‘I’ and ‘you’ (sg.) are assembled for as many languages as possible (about 120 in all<sup>3</sup>), with a certain degree of latitude about semantic resemblance (also some plural forms are accepted) and word class status (nominal and verbal affixes are also included), with the implicit hope that the more languages that are involved in the comparison, the more likely it is that a distinctive pattern of resemblance unrelated to chance will emerge. Trombetti’s indication of a widespread *ni* : *m* pronominal pattern in the New World does not seem, however, to have reached the linguists who, in the early 1900s, in North America, dedicated themselves to the description and classification of the languages of that continent. The following observations by Roland B. Dixon, Alfred L. Kroeber and Edward Sapir seem to indicate that these authors were aware of the wide distribution of the *n* : *m* pronominal set independently from Trombetti.

“It will be seen that as in so many American languages, the pronominal stems of the first and second person are based on *n* and *m*.” (Dixon 1910: 322)

“Throughout the field of linguistic structure in the whole continent, there are abundant examples of the operation of the principle of territorial continuity of characteristics, and of the underlying one that even the most diverse languages affect each other, and tend to assimilate in form, if only contact between them is intimate and prolonged. Such are the exceedingly common occurrence [sic] of *n* and *m* to designate the first and second person pronouns; the geographical localization of families expressing sex gender; the prevailing tendency for pronominal elements, especially the possessive ones, and instrumental elements in verbs, to be prefixes rather than suffixes, as already mentioned for California. It is needless to multiply examples which are either familiar to the Americanist or readily compilable by him.” (Kroeber 1913: 399)

“The curiously widespread American second person singular in *m*-meets us here [in Mosetén, a language spoken in Bolivia] once more (*mi* “thou”).” (Sapir 1918: 184)

“Getting down to brass tacks, how in the Hell are you going to explain general American *n*- “I” except genetically? It’s disturbing to know but (more) non-committal conservatism is only dodging after all, isn’t it?” (Letter from Sapir to Frank G. Speck in 1918; cited in Campbell 1994: 2)

The first scholar to explicitly attribute the *n* : *m* pronominal pattern to specific American families, rather than to single languages, was Morris Swadesh in the 1950s.

“At least two short elements, *n* for the first person pronoun and *m* for the second (...) are so numerous as to virtually eliminate the chance factor despite their brevity. In fact, even if one disregarded the cases which have one or the other and included only the languages which have both *n* and *m* for first and second person respectively, and if one holds to the restriction that both forms must belong to the same functional type—whether independent pronoun or subject, object or possessive affixes—the list of language groups would still be fairly impressive. It would include families of the Penutian and Hokan-Coahuiltecan phyla, Aztecanoan, Chibchan, and Mapuche.” (Swadesh 1954: 311-312)

Swadesh was also the first scholar to assign the *n* : *m* set to a common ancestor of all, or at least most, Native American languages. For “Proto Ancient American”, in 1960, he proposed *\*(?e)ne*, often *\*(?i)ni*, *T* and *\*ma/\*mu* ‘thou, thy’ (Swadesh 1960: 907-909). Similar reconstructions appeared in 1972 in Esther Matteson’s first attempt to create a large-scale reconstruction of “Proto-Amerindian”. Among the 974 Proto-Amerindian forms reconstructed by Matteson (1972), we find *\*nV T* (p. 65) and *\*mV* ‘you’ (sg.) (p. 89).

In the 1970s, the geographical distribution of the *n*- and *m*-pronominal forms in the Americas was still unclear, however. In an article of 1976 on northern California as a linguistic area, Mary H. Haas wrote as follows.

“The most prominent feature is *n*- in the first person paired with *m*- in the second person. This is found in Karok (Hokan) and in Klamath, Maidu, Wintu, and Yokuts (Penutian) (...)  
But the total picture of diffusion of *n*- and *m*- in the first and second persons goes beyond the area being studied in this paper and so the problem really needs to be attacked on a larger scale. The use of *n*- in the first person (paired with other things in the second person as well as with *m*-) is also very widespread and in this wider context the first person *n*- in Algonkian-Wiyot-Yurok is also part of the picture.” (Haas 1976: 358-359)

Many of Matteson’s (1972) etymologies coincide with those included by Joseph Greenberg in the volume *Languages in the America* of 1987 in which, through a “multilateral” comparison of grammatical morphemes and lexical forms of a large portion of the

languages of the New World, the latter author attempts to demonstrate the genealogical unity of all American languages with the exception of Eskimo-Aleut and Na-Dene (Athabaskan-Eyak-Tlingit and Haida) of the extreme north of North America. Among the some 2,000 “Amerind” etymologies assembled by Greenberg, the *n* : *m* pronominal pair could not be omitted (pp. 48-54). Greenberg regards the two pronominal forms as one of the stronger pieces of evidence for his Amerind macrogrouping: “[i]n fact, it would probably be easier to enumerate where *nV*- and *mV*- are not found [among American language families] than where they are”, he wrote (p. 49). And his collection of *n*- and *m*-forms involves a number of languages greater than that reached by Trombetti: about 150, in addition to some proto-languages. The *n*- and *m*-forms, as in Trombetti (1905), are personal pronouns and pronominal affixes, both singular and, not rarely, plural.

In 1994, Lyle Campbell and, in 1996, Johanna Nichols and David A. Peterson argued against Greenberg’s claim that the American *n* : *m* pattern reflects a pattern in the common ancestor of “Amerind” languages.

Campbell (1994) makes us note that the pattern is neither pan-Amerind nor distinctive of Amerind, citing various nonconforming pronominal systems from American languages and various conforming ones from elsewhere in the world (pp. 6-8). He also points out that nasals are to be expected in pronouns due to their perceptual salience (pp. 3-4) and/or for having indirectly to do with nursing infants (p. 6), that pronominal forms can be borrowed (pp. 4-5), and hence that the use of pronominals with nasals to establish genealogical relationships between languages is tricky at best.

By examining “a moderately large sample of the world’s language families”, including 71 American languages and another 102 languages spoken in the rest of the world, “and a strict grid of pronominal forms and categories for consistency of comparison”, Nichols & Peterson (1996) propose that the *n* : *m* paradigm has what they call a “Pacific Rim” distribution, rather than being a specifically pan-American phenomenon, and that this distribution is not limited to America, but also includes New Guinea.

“The *n* : *m* paradigm has a geographical distribution similar to that of *m* by itself: it is found chiefly in western North America, Mesoamerica, and western South America, and marginally in northern coastal New Guinea. (...)”

Within both New Guinea and the New World there are again coherent geographical distributions. Languages showing the paradigm, in either loose or strict form, are found in New Guinea only on the

northern coast and northern offshore islands and in the Americas only in the west. (...)

We call this distribution—northern coastal New Guinea-Melanesia and the western New World—the Pacific Rim distribution.” (Nichols & Peterson 1996: 361)

Based on the assumption that the overall probability of occurrence of the *n* : *m* paradigm is fairly low on statistical grounds (cf. also Nichols 1996: 55), Nichols and Peterson also indicate that the paradigm, together with other Pacific Rim features like tones and numeral classifiers, is an excellent “marker” of a single, non-initial phase in the settlement of the Americas.

“Given that the Americas were settled by immigration through Beringia, the Pacific Rim features define what appears to be a relatively recent phase in colonization. At least since the end of glaciation, when sea levels rose to their present highs, moving from Siberia to Alaska has involved crossing open ocean, an option available only to coastally adapted people familiar with watercraft. The present distribution of Pacific Rim features further suggests that immigrants retained their coastal orientation long after entry. Pacific Rim features have spread over ten thousand miles down the Pacific coast but at most only about one thousand miles inland and that only sporadically.” (Nichols & Peterson 1996: 368)

In response to Nichols and Peterson, Campbell (1997a) argued that what the two authors have shown is due to nothing more than chance and highlights, *inter alia*, two objective weaknesses of their survey: an inadequate presence of American languages in the sample used and an inattention to the history of the languages under consideration or, at least, of their *n*- and *m*-forms that poorly goes with the value of the historical marker attributed to the *n* : *m* paradigm. With regard to the first point, Campbell makes us observe that, although Nichols and Peterson’s sample is highly skewed towards America, the 71 Native American languages it includes are not fully representative of the approximately 150 unrelated language families that the New World contains (p. 342). With regard to the second point, Campbell makes us aware that the assumption of continuity of the *n* : *m* paradigm over time and within language families is a precarious position to take because a common sound change like, for example, final *-m* to *-n* occurred in the Balto-Finnic languages, that resulted in first-person singular *-m* (of Nichols and Peterson’s Old World pattern; cf. Nichols & Peterson 1996: 360) being converted into *-n* (closer to their Pacific

Rim pattern), “potentially could render a major shift in a language’s (or even a whole language family’s position) in Nichols and Peterson’s scheme” (pp. 344-344). He signals, in addition, that at least two of the nineteen South American languages of Nichols and Peterson’s sample, namely Axininca Campa (an Arawakan language) and Hixkaryana (a Cariban language), owe their current pronominal systems to local innovations and do not belong in an  $n : m$  set when their history is taken into account, with the consequence that 11% of the South American representatives have been misplaced (p. 344).

We should also note that a strictly synchronic survey of modern languages, like that of Nichols and Peterson, cannot even consider the fact that, with time, nasals may be subject to denasalization or, also, to loss. Accordingly, languages that today do not have the  $n : m$  paradigm, but did have it historically, end up being completely neglected by this type of survey. In section 1 of Appendix B, we may note that it is due to internal evidence that the Wiyot first-person prefix *d-* can be considered as a reflex of Proto-Algic  $*n-$  ‘1’ (but cf. also Southern Puget Salish and Twana *d-* ‘1sg’ (possessive prefix) from Proto-Salishan  $*n=$  ‘id.’ (Newman 1979: 211)). In Terena, the possessor/subject prefix  $*nu-$  ‘1sg’ of Proto-Arawakan has become a process of spreading of nasalization from the left edge of the word which yields nasal vowels and/or prenasalized consonants (Carvalho 2017: 517, 518, 524-529), while Chamicuro, another member of the Arawakan family, has innovated  $u- \sim w-$  for ‘1sg’ (Aikhenvald 1999: 88). In Kuikuro, the second-person singular personal pronoun, ultimately from Proto-Cariban  $*VmV$  (perhaps  $*amo$ ), has an unexpected form  $e:ye$  with an intervocalic  $m$ -loss: Pre-Kuikuro  $*eme-ye > e:ye$  (Meira 2002: 260). The second-person singular personal pronoun  $a:pi$  ‘you’ (sg.) of Lower Pima comes from Proto-Uto-Aztecan  $*i(mi)$  ‘id.’ and similar examples of first or second-person (singular) pronominal forms that do not retain the nasal feature of the  $n$  or  $m$  present in their etymon are not uncommon to find among the Native languages of the Americas. Forms of modern languages, in short, do not offer the best observation post of a phenomenon which is considered as being of very ancient origin like the American  $n : m$  paradigm is.

In 1998, Nichols and Peterson replied to Campbell’s observations restating that they “showed clearly that the  $n : m$  system’s distribution cannot simply be ascribed to chance” and that “[s]omething happened”, although they “do not and cannot know just what happened” (Nichols & Peterson 1998: 613).

A rather recent contribution by Nichols and Peterson shows on the map the geographical distribution of the  $n : m$  paradigms

on a global scale based on a sample including 90 languages from Native America and further 140 languages from the rest of the world (Nichols & Peterson 2005<sup>4</sup>). The paradigms appears to form a large areal cluster extending from western coastal North America to western South America (Map 137). The authors' conclusion is that the American *n : m* paradigms are the product "of geographical spread rather than just universals or just inheritance" and that their origins "are old (older than any of the individual families exhibiting the systems, for instance)": "late glacial at the latest" (p. 549).

Most (67) of the American languages of Nichols and Peterson's 2005 sample are also present in their 1996 sample; nine (31%) of the 29 American *n : m* paradigms mapped are of secondary origin, i.e. not inherited from the proto-languages of the families to which the languages in which they occur belong.<sup>5</sup>

## *2. Old and new data*

Two decades since the contributions of Nichols & Peterson (1996, 1998) and Campbell (1997a), I believe that it is opportune that we return to deal with the *n : m* American pronominal paradigm if only to have an overview, as precise as possible, of its frequency and geographical spreading in the New World. Thanks to a significant number of recent grammars and dictionaries of languages spoken in South America of which, twenty years ago, little or nothing was known, we can today define with (relative) precision the territorial distribution of the pronominal *n*- and *m*-forms in this continent. We moreover now have a better knowledge of the genealogical relationships of many languages, both of North and South America. Because of the impossibility of taking into account every single attested American language, this knowledge is of fundamental importance to constitute a collection of *n*- and *m*-forms on the basis of a sample of languages fully representative of the extraordinarily large range of linguistic variation that characterizes the Americas.

Sections 1 and 2 of Appendix B to the present article present an inventory of American *n*- and *m*-pronominal forms for first and second person gathered on the basis of a linguistic sample that is as balanced as possible. Section 3 surveys the geographical distribution of these forms in the entire New World, while section 7 shows their frequency in specific areas. A list of the *n* forms occurring together with an *m*-form is given in section 4. The remaining sections of this article deal with the American languages that have neither *n* nor *m*

in first-person or second-person pronominal forms (section 5 and sections 3-6 of Appendix B) and the diffusion of first-person singular *n* and second-person singular *m* in Eurasia (section 6 and sections 7 and 8 of Appendix B).

Regarding the pronominal forms listed in Appendix B, these few clarifications are necessary.

- (i) By *n*, I intend dental or alveolar [n], palatal [ɲ], or alveopalatal [ɳ] (though very few pronominal forms with [ɲ] or [ɳ] occur in my sample). *m* is a bilabial.
- (ii) The pronominal forms are independent pronouns and bound personal designators (clitics and affixes) for first person singular, second person singular, or expressing first person or second person without specification for number.<sup>6</sup> To limit the possibilities of chance resemblances, the only pronominal forms of first person plural and second person plural with *n* or *m* included in Appendix B are those derived from or etymologically related to the corresponding singular forms (like Proto-Tucanoan \**mĩĩ-sã* ‘you’ (pl.) from \**mĩĩ* ‘you’ (sg.)).
- (iii) Only pronominal forms in which *n* or *m* is the first consonant were picked up. This constraint permits strict comparability of forms. Some pronominal forms in which *n* or *m* is the second consonant are however shown in Appendix B since they appear clearly connected with coreferential forms of the same language or proto-language with *n* or *m* as first consonant (e.g. the habitual ending *-ka’ni* ‘1sg’ of Tunica, that appears constructed from the semelfactive ending *-ni* ‘1sg’ by the addition of an initial element *-ka-*).
- (iv) In the case of forms which are analyzable synchronically or are, at least, diachronically segmentable in two or more constituents, the first *n* or *m* is, of course, part of a pronominal affix, rather than of a recurring element (as in Proto-Yokutsan \**na-ʔ* ‘I’ and \**ma-ʔ* ‘you’ (sg.), not as in Guamo <*na-pi*> ‘I’ and <*na-jâ*> ‘you’ (sg.)).

In order to account for the extraordinary linguistic variation that characterizes both North and South America, the method I adopted to constitute a sample that adequately represents the two continents is the one that requires that a sample must (ideally) contain one representative from each independent family (Rijkoff & Bakker 1998: 268). In doing that, I treated language isolates as single families, with the consequence that all known American language isolates are included in my sample. For families with internal structure, the representative

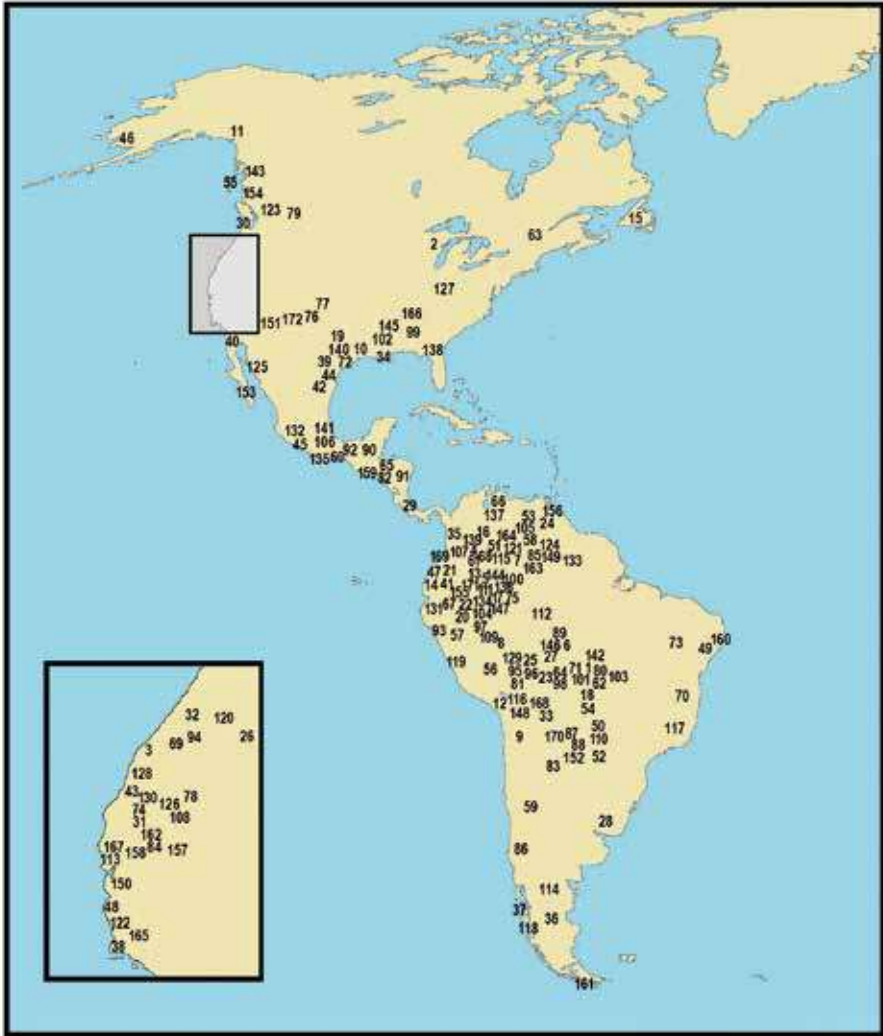


member I chose, where possible, is the top node, the proto-language, not a modern descendant. We today have (more or less detailed) reconstructions of the proto-languages of most of the known American multi-member language families (see Table 1 in Appendix A), but not all. For the families that have not yet been the object of comparative-reconstructive study, I gathered forms of, at least, one representative language. In general, these forms should not be too dissimilar from the corresponding form of the proto-language of the family.

Obviously, the choice of basing the examination of the distribution in the Americas of the  $n : m$  pronominal paradigm on a sample which is largely formed of reconstructed languages implies some degree of risk, i.e. that some of the proto-languages under consideration are not a scientific product with a firm consensus among expert linguists and/or present critical issues like, for example, not having been reconstructed based on all the forms of the descendant languages that it would have been opportune to examine. On the other hand, a sample that takes into account only proto-languages of families and language isolates permits the risk to be excluded that the collection of  $n$ - and  $m$ -pronominal forms obtained through it includes elements that (at the present state of our knowledge) demonstrably descend from old pronominal forms not containing a nasal consonant.

There is yet a wide margin of disagreement among linguists as to how many and which families (including isolates) should be recognized in the Americas. The long Table 2 in Appendix A, based (with a rather wide margin of freedom) on *Glottolog 3.0* (Hammarström, Forkel & Haspelmath 2017), provides a list of 172 generally accepted language families, including 89 families that have at least two languages as members and 83 language isolates. For each family the following information is provided: (i) Glottocode, (ii) geographical distribution, and (iii) number of languages. Map 1 shows the approximate location of the proto-languages of the 89 families with at least two languages and of the 83 isolates.

It has to be noted that, in this map, I located the proto-languages of multi-member families with (relatively) small territorial extent near the center of their territory, and that I placed the proto-languages of families with great territorial extent, with the exception of Nuclear Macro-Je, near those that are considered to be their Urheimaten. The families for which I maintain that we can speak of great territorial extent, in addition to Nuclear Macro-Je, are fifteen. Their supposed Urheimaten are indicated in Table 3 in Appendix A.<sup>7</sup> The literature I consulted does not contain claims about the homeland of the Nuclear Macro-Je family, as defined in *Glottolog 3.0*, and



**Map 1.** Location of the proto-languages of the linguistic families and of the language isolates of the Americas. (Proto-languages and language isolates are identified by the same numbers in Table 2 in Appendix A.)

the location of this family in Map 1 is simply a central one compared to the geographical distribution of its known languages. Finally, in identifying the location of isolates, I used their locations at the time of European contact.

For the sake of completeness, I also examined the very poor documentation of all American unclassified languages of which I am aware of in search of further *n*- and *m*-pronominal forms. These are 66 extinct languages of which all that remain are only a few words (excluding proper names), just one or two in some cases: too few to attempt to classify them (see Table 4 in Appendix A). No *n*-form for '1(SG)' nor *m*-form for '2(PL)' came to the surface from this examination.

### *2.1. The n-forms*

First-person or first-person singular *n*-forms occur in the pronominal system of 28 proto-languages and 22 isolates (i.e. 50 families) of my sample. A list of the *n*-forms is provided in section 1 of Appendix B.

- (1) NORTH AMERICA: Cayuse, Chimariko, Coahuilteco, Esselen, Karankawa, Karok, Klamath-Modoc, Molala, Natchez, Proto-Algic, Proto-Chinookan, Proto-Chumashan, Proto-Cochimí-Yuman, Proto-Huavean, Proto-Jicaquean, Proto-Keresan, Proto-Kiowa-Tanoan, Proto-Lencan, Proto-Maiduan, Proto-Mayan, Proto-Mixe-Zoquean, Proto-Sahaptian, Proto-Salishan, Proto-Tsimshianic, Proto-Uto-Aztecan, Proto-Wakashan, Proto-Wintuan, Proto-Xincan, Proto-Yokutsan, Siuslaw, Takelma, Tarascan, Tunica.
- (2) NORTH & SOUTH AMERICA: Proto-Arawakan.
- (3) SOUTH AMERICA: Arara of Rio Branco, Cofán, Mapuche, Muniche, Omurano, Proto-Andoque-Urekena, Proto-Aymaran, Proto-Guahiboan, Proto-Huarpean, Proto-Otomacoan, Proto-Quechuan, Proto-Tupian, Puelche, Puquina, Urarina, Warao.

The attribution of an *n*-form for '1(SG)' to the proto-languages of nine families is, for various reasons, uncertain (see again section 1 of Appendix B).

- (1) NORTH AMERICA: Proto-Alsean, Proto-Coosan, Proto-Otomanguan, Proto-Timucuan, Proto-Utian.
- (2) NORTH & SOUTH AMERICA: Proto-Chibchan.
- (3) SOUTH AMERICA: Proto-Candoshi-Chirino, Proto-Timotean, Proto-Uru-Chipaya.

There are not (recognizable) first-person (singular) pronominal forms in the scanty documentation of the languages of the extinct Tallán and Xukuruan families of South America and of the also extinct Beothuk and Cotoname of North America and Chono of South America.

## 2.2. *The m-forms*

Second-person or second-person singular *m*-forms occur in the pronominal system of 35 proto-languages and 23 isolates of my sample (viz. 58 families).<sup>8</sup> A list of the *m*-forms is provided in section 2 of Appendix B.

- (1) NORTH AMERICA: Cayuse, Chimariko, Coahuilteco, Karankawa, Karok, Klamath-Modoc, Molala, Proto-Chinookan, Proto-Cochimí-Yuman, Proto-Huavean, Proto-Kalapuyan, Proto-Kiowa-Tanoan, Proto-Lencan, Proto-Maiduan, Proto-Misumalpan, Proto-Mixe-Zoquean, Proto-Palaihnihan, Proto-Pomoan, Proto-Sahaptian, Proto-Tequistlatecan, Proto-Totonacan, Proto-Tsimshianic, Proto-Utian, Proto-Uto-Aztecan, Proto-Wintuan, Proto-Yokutsan, Proto-Yukian, Salinan, Seri, Takelma, Washo.
- (2) SOUTH AMERICA: Aikanã, Kanoë, Leco, Mapuche, Mosetén-Chimane, Mure, Proto-Aymaran, Proto-Cariban, Proto-Chapakuran, Proto-Charruan, Proto-Chonan, Proto-Guahiboan, Proto-Jivaroan, Proto-Kakua-Nukak, Proto-Nadahup, Proto-Nambiquaran, Proto-Panoan, Proto-Tacanan, Proto-Tucanoan, Proto-Uru-Chipaya, Proto-Zamucoan, Puelche, Puinave, Vilela, Waorani, Yaruro, Yurakaré.

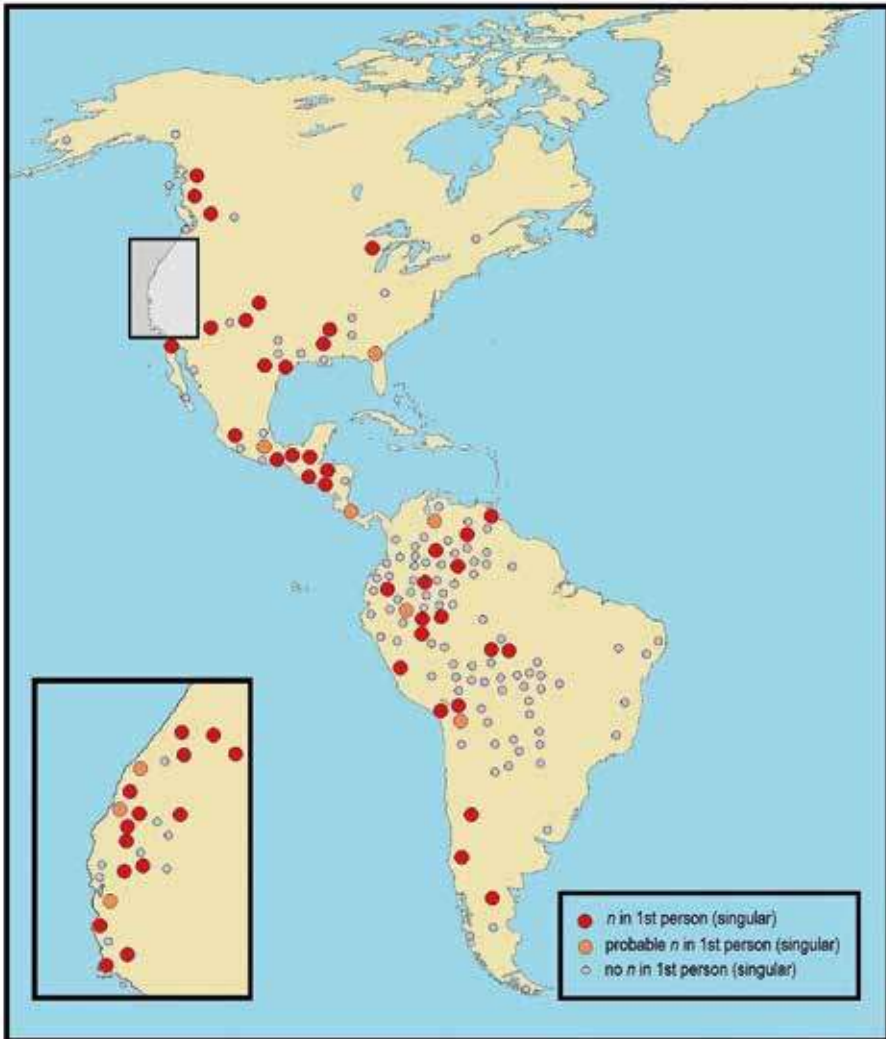
In some cases, it is impossible to establish with certainty whether a proto-language has an *m*-form expressing second person (singular) (see again section 2 of Appendix B).<sup>9</sup>

- (1) NORTH AMERICA: Proto-Shastan.
- (2) NORTH & SOUTH AMERICA: Proto-Chibchan.
- (3) SOUTH AMERICA: Proto-Híbito-Cholón, Proto-Jirajaran, Proto-Timotean.

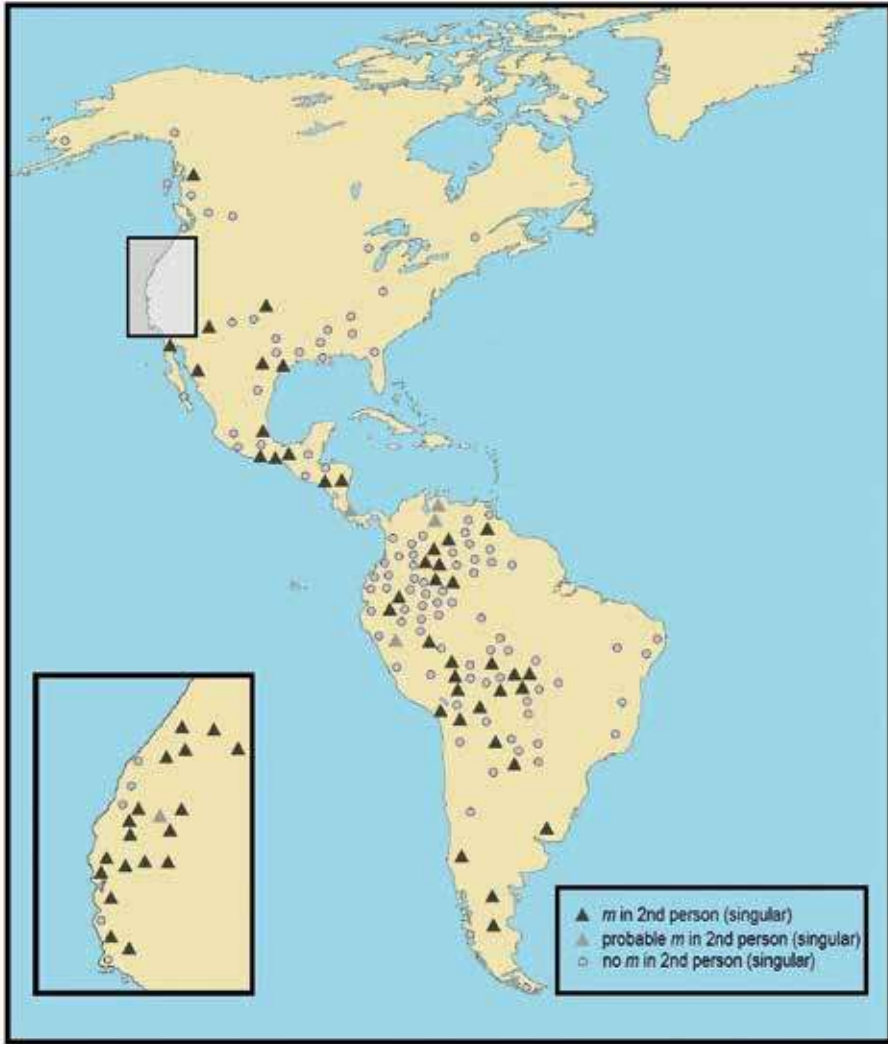
There are not (recognizable) second-person (singular) pronominal forms in the scanty documentation of the languages of the extinct Tallán and Xukuruan families of South America and of the also extinct Beothuk and Cotoname of North America and Arara of Rio Branco, Chono, Guachí, and Omurano of South America.

*3. Geographical distribution of n- and m-forms*

Map 2 shows the geographical distribution of the *n*-forms for '1(SG)' listed in section 1 of Appendix B, while Map 3 shows that of *m*-forms for '2(SG)' listed in section 2 of the same appendix. In Map 4, we may observe the joined geographical distribution of the *n*- and *m*-forms. Twenty-four asterisks represent the proto-languages or lan-



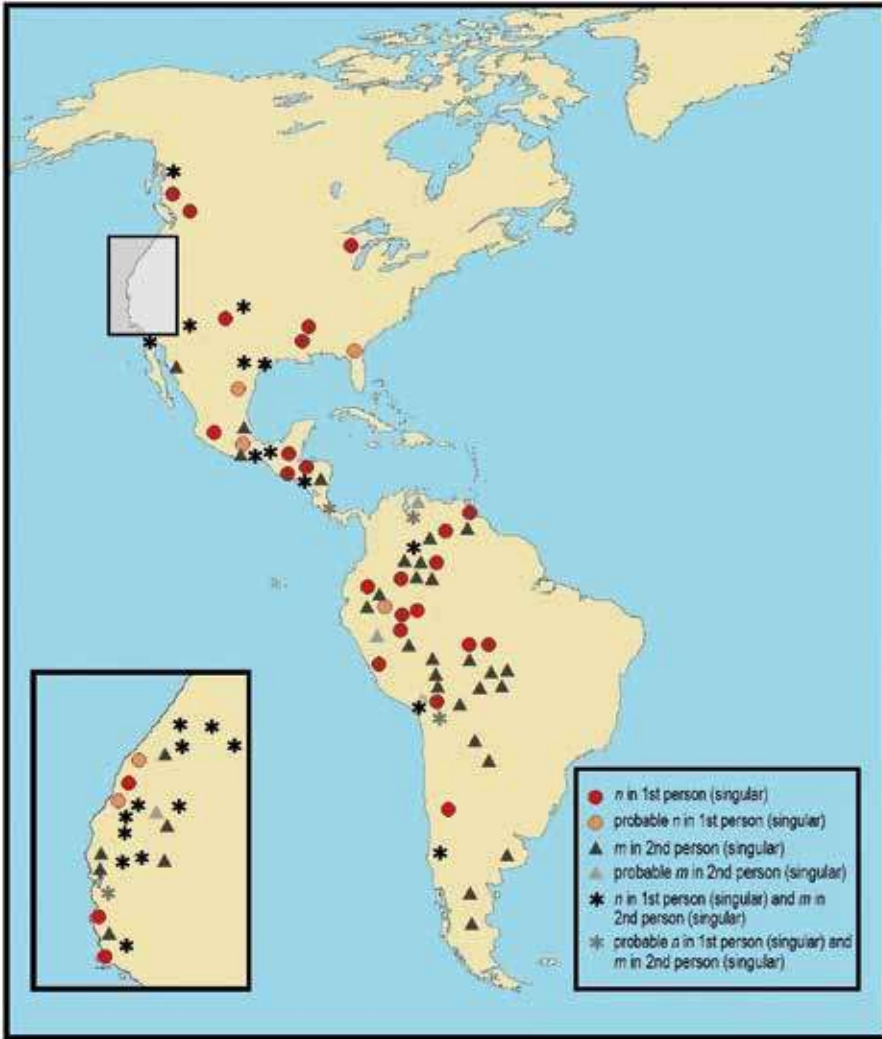
**Map 2.** *n* in first person (singular) (Americas).



**Map 3.** *m* in second person (singular) (Americas).

guage isolates in which, based on the literature consulted, the *n* : *m* paradigm occurs (see section 4 for their list).

It should be kept in mind that these three maps give a panchronic representation of the geographical distribution of the *n*- and *m*-forms, not a synchronic one restricted to the present or to a specific time in the past. The same *n*- and *m*-forms belong to language isolates



**Map 4.** *n* in first person (singular) and/or *m* in second person (singular) (Americas).

that are still spoken or that ceased to be spoken in different periods as well as to proto-languages that have different levels of time depth like, for example, Uto-Aztecan (4,000 or 5,000 years) and Chumashan (perhaps no more than 1,000 years).

Map 2 highlights a strong presence of *n*-forms in North America and a weaker presence in South America. Map 3 highlights a large

presence of *m*-forms both in North and South America. In regard to North America, although the majority of the *m*- and *n*-forms is concentrated in the Pacific Rim region, we can note that the remaining *m*- and *n*-forms have a rather homogeneous distribution in the rest of the continent. In South America, the *n*- and *m*-forms are mainly distributed in the western half, but only to a very limited extent on the Pacific Rim area.

We should also keep in mind that what Maps 2-4 show is the distribution of *n*- and *m*-forms and proto-forms in the New World that we are today able to reconstruct based on our knowledge of its Native languages. Because there are four large areas in the Americas whose linguistic prehistories we know very little of, this reconstruction should be regarded as being absolutely partial.

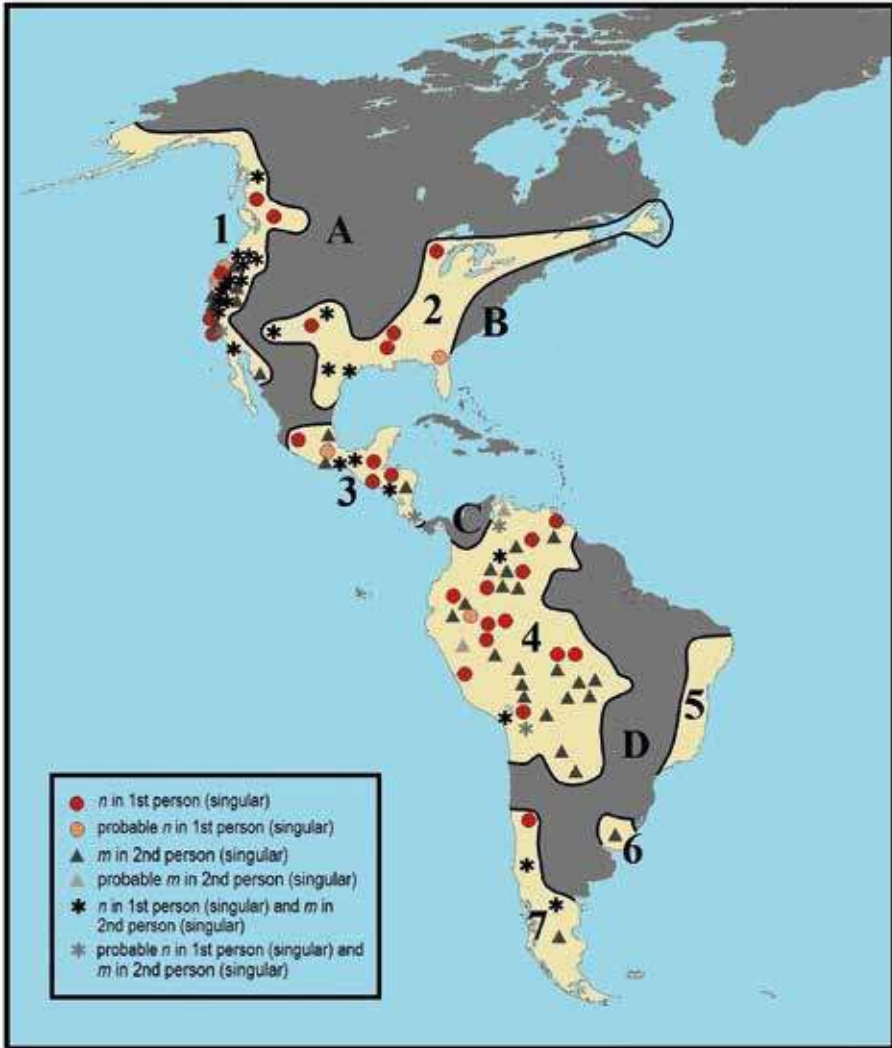
The northern and more extended of these four large areas is the one identified by the letter A in Map 5. At the time of contact, this area was dominated by languages of the following five families: Eskimo-Aleut, Athabaskan-Eyak-Tlingit, Algic, Siouan-Catawba, and Uto-Aztecan. None of the supposed Urheimaten of these families, however, lies inside area A, and we do not know any language isolate that was spoken there, although we have some modest testimony of extinct unclassified languages once spoken in the southern extremity (Concho, Jumano-Suma, Maratino, Naolan, Quinigua, Solano, and Tanpachoa; Table 4 in Appendix A).

A second area, identified by the letter B in Map 5, is the region between the Appalachian Mountains and the Atlantic. The languages spoken here at the time of contact belong to the Algic, Siouan-Catawba, and Iroquoian families and include two (now extinct) unclassified languages (Nansemond and Pamunkey; Table 4 in Appendix A).

A third area, which is smaller than the preceding ones and is identified by the letter C in Map 5, includes the present-day state of Panama and the northwestern sector of Colombia. No known language family has its Urheimat in this area and no isolate language is known to have been spoken here. Just some extinct unclassified languages are from area C (Colima, Cueva, Idabaez, Malibú, Muzo, Panche, Pijao, Quimbaya, Tairona, and Yamesí; Table 4 in Appendix A).

The fourth area is the one identified with the letter D in Map 5. This area covers the eastern portion of the Guiana Highlands, eastern Amazonia, the northern section of the Brazilian Highlands, the southern portion of the Gran Chaco, the Pampas (excluding an area in the east where the languages of the Charruan family were once spoken), the territory between the Pampas and the Brazilian Highlands, and a





**Map 5.** *n* in first person (singular) and/or *m* in second person (singular) and the areas of the Americas whose linguistic prehistory is unknown.

sector of the Andes west of the Gran Chaco. At the time of contact, the Guiana Highlands, eastern Amazonia, and the northern section of the Brazilian Highlands area were occupied by groups speaking Cariban, Arawakan, Tupian, and Nuclear Macro-Je languages as well as several other groups whose languages we know very little (Baenan, Gamela

of Viana, Kaimbé, Kambiwá, Natú, Pankararú, Tarairiú, Tuxá, and Xocó; see Table 4 in Appendix A) or nothing of. The rest of this area is practically a *terra incognita* from a linguistic point of view. At the time of contact, it was in great part occupied by groups that have now disappeared and whose languages nothing or almost nothing remains (Boreal Pehuelche, Comechingón, Guachipas, Diaguita, Otí, Querandí, Sanavirón, and Tembey; Table 4 in Appendix A).

If we exclude these four areas from our overview of the New World, we can observe seven non-contiguous areas in which *n* and *m* are present in a somewhat diversified manner. The seven areas, indicated by numbers in Map 5, are the following: 1. North American Pacific Rim; 2. Great Plains, Southern Plains, Southeast United States, Midwest United States, Saint Lawrence Lowland, and Newfoundland; 3. Mesoamerica and Central American Isthmus; 4. Central and northern sectors of western South America; 5. São Francisco basin and eastern and southeast Atlantic regions of Brazil; 6. Eastern Pampas; 7. Southern Andes and southern Patagonia.

Both *n*- and *m*-forms are completely absent in area 5. In area 6, we only find the *m*-form of Proto-Charruan. Many proto-languages and isolates of area 1 have the complete *n* : *m* paradigm. In areas 2 and 3, some proto-languages and isolates that have the *n* : *m* paradigm are to be found near a greater number of proto-languages and isolates that have only one of the components of the paradigm. In area 4, we may note a substantial number of *m*-form and a more limited number of *n*-forms. The probability that one of the frequent *m* co-occurs with *n* is reasonably high, but actually, only two or, perhaps, at most four proto-languages in the wide area 4 have the complete *n* : *m* paradigm.

The high concentration of *n*- and *m*-forms in the southern half of area 1, California, goes hand in hand with the striking density of multi-member language families and isolates of this zone that, although usually considered as independent, might belong to wider groupings. If the genealogical validity of such groupings will be convincingly demonstrated by future research, we would have a lesser number of families in the southern half of area 1 whose proto-languages would presumably attest yet older *n*- and *m*-forms.

#### 4. Languages with both *n*- and *m*-forms

Fourteen proto-languages and ten language isolates of my sample have (at least) both one *n*-form and one *m*-form. I thought it would be

*First-person n and second-person m in Native America: a fresh look*

useful to gather most of these forms here below. (Proto-forms preceded by two asterisks are my tentative reconstructions based on the forms of the modern languages shown in sections 1 and 2 of Appendix B.)

|                               |   |   |
|-------------------------------|---|---|
| NORTH AMERICAN<br>PACIFIC RIM | Cayuse                                    | 1SG <i>íniŋ</i> (personal pronoun)                  |
|                               |   | 1SG <i>nīs-</i> (S prefix)                          |
|                               |   | 2SG <i>mīs-</i> (S prefix)                          |
|                               | Chimariko                                 | 1SG <i>noʔot</i> (personal pronoun)                 |
|                               |   | 2SG <i>mamot</i> (personal pronoun)                 |
|                               |   | 2SG <i>m-</i> (possessive/S/O prefix)               |
|                               |   | 2SG <i>-m</i> (possessive/S/O prefix)               |
|                               | Karok                                     | 1SG <i>ná·</i> (personal pronoun)                   |
|                               |   | 1SG <i>nani-</i> ~ <i>nini-</i> (possessive prefix) |
|                               |   | 1SG>3 <i>ni-</i> ~ <i>ná-</i> (S-O prefix)          |
|                               |   | 2SG <i>ʔí·m</i> (personal pronoun)                  |
|                               |   | 2SG <i>mi-</i> (possessive prefix)                  |
|                               | Klamath-Modoc                             | 1SG <i>ni-s</i> (obj.) (personal pronoun)           |
|                               |   | 2SG <i>mi-s</i> (obj.) (personal pronoun)           |
|                               | Molala                                    | 1SG <i>=in</i> (possessive enclitic)                |
|                               | 1SG <i>n-</i> (O prefix)                  |   |
|                               | 2SG <i>=im</i> (possessive enclitic)      |   |
|                               | 2SG <i>m-</i> (O prefix)                  |   |
| Proto-Chinookan               | 1SG <i>**n-aiyak/x</i> (personal pronoun) |   |
|                               | 1SG <i>**n-</i> (S/O/possessive prefix)   |   |
|                               | 2SG <i>**m-aiyak/x</i> (personal pronoun) |   |
|                               | 2SG <i>**m-</i> (S/O/possessive prefix)   |   |
| Proto-Cochimí-Yuman           | 1SG <i>**n-</i> (S prefix)                |   |
|                               | 2SG <i>**m-</i> (S prefix)                |   |
| Proto-Maiduan                 | 1SG <i>*ni</i> (personal pronoun)         |   |
|                               | 2SG <i>*mi</i> (personal pronoun)         |   |
| Proto-Sahaptian               | 1SG <i>*ʔí-n</i> (personal pronoun)       |   |
|                               | 2SG <i>*ʔí-m</i> (personal pronoun)       |   |
| Proto-Tsimshianic             | 1SG <i>*=nə</i> (S enclitic)              |   |
|                               | 2SG <i>*=mə</i> (S enclitic)              |   |
| Proto-Wintuan                 | 1SG <i>*ni</i> (personal pronoun)         |   |
|                               | 2SG <i>*mi</i> (personal pronoun)         |   |

|  |                    |  |
|--|--------------------|--|
|  | Proto-Yokutsan     | 1sg * <i>na-ʔ</i> (personal pronoun)<br>2sg * <i>ma-ʔ</i> (personal pronoun)   |
|  | Takelma            | 1sg - <i>àn</i> ~ - <i>n</i> (S suffix)<br>2sg <i>ma</i> ~ <i>ma:</i> (personal pronoun)<br>2sg, 2sg>1sg - <i>tam</i> (S/S-O suffix)<br>2sg>1sg - <i>kaʔm</i> (S-O suffix)   |
| GREAT PLAINS,<br>SOUTHERN PLAINS,<br>SOUTHEAST US,<br>MIDWEST US,<br>SAINT LAWRENCE<br>LOWLAND &<br>NEWFOUNDLAND | Coahuilteco        | 1sg <i>n-ami-</i> (possessive pronoun)<br>1sg <i>n-</i> ~ <i>na-</i> (possessive prefix)<br>1sg>2/3PL <i>nak-</i> (S-O prefix)<br>1sg(>3) <i>niw-</i> (S/S-O prefix)<br>2sg>3PL, 3>2 <i>mak-</i> (S-O prefix)<br>2sg(>3) <i>may-</i> ~ <i>mi-</i> (S/S-O prefix) |
|  | Karankawa          | 1sg <i>náyi</i> (personal pronoun)<br>1sg <i>n-</i> (S prefix)<br>2sg <i>m-</i> (S prefix)   |
|  | Proto-Kiowa-Tanoan | 1sg * <i>nə</i> (personal pronoun)<br>2sg * <i>wim</i> (personal pronoun)  |
|  | Proto-Uto-Aztecan  | 1sg *( <i>i</i> ) <i>ni</i> (personal pronoun)<br>1sg *( <i>i</i> ) <i>ni-</i> (possessive prefix)<br>2sg * <i>i</i> ( <i>mi</i> ) (personal pronoun)<br>2sg * <i>i</i> ( <i>mi</i> )- (possessive prefix)   |
| MESOAMERICA &<br>CENTRAL AMERICAN<br>ISTHMUS   | Proto-Huavean      | 1sg * <i>nV-</i> (S prefix)<br>1sg * <i>-nV</i> (S suffix)<br>2sg * <i>m-</i> (S prefix)<br>2sg * <i>-m</i> (S suffix)   |
|  | Proto-Lencan       | 1sg ** <i>-on</i> ~ ** <i>un</i> (possessive suffix)<br>2sg * <i>ama-nani</i> (personal pronoun)<br>2sg ** <i>am-/ma-</i> (possessive prefix)<br>2sg ** <i>-mi</i> (S suffix)  |
|  | Proto-Mixe-Zoquean | 1sg * <i>n-he?</i> (possessive pronoun)<br>2sg * <i>mici</i> (personal pronoun)<br>2, 2>3, 3>2 * <i>min-</i> (S-O prefix)  |

*First-person n and second-person m in Native America: a fresh look*

|  |                 |  |
|--|-----------------|--|
| CENTRAL &<br>NORTHERN SECTORS<br>OF WESTERN SOUTH<br>AMERICA | Proto-Guahiboan | 1SG * <i>xá-ni</i> (personal pronoun)    |
|  |                 | 2SG * <i>xá-mi</i> (personal pronoun)    |
|  | Proto-Aymaran   | 1SG * <i>na-ya</i> (personal pronoun)    |
|  |                 | 1SG *- <i>Na</i> (possessive suffix)     |
|  |                 | 2SG * <i>hu-ma</i> (personal pronoun)    |
|  |                 | 2SG *- <i>ma</i> (S/possessive suffix)   |
| SOUTHERN ANDES<br>& SOUTHERN<br>PATAGONIA                    | Mapuche         | 1SG <i>ijñé</i> (personal pronoun)       |
|  |                 | 1SG - <i>n</i> ~ - <i>in</i> (S suffix)  |
|  |                 | 2SG <i>eymi</i> (personal pronoun)       |
|  |                 | 2SG <i>mi</i> (possessive pronoun)       |
|  |                 | 2SG - <i>m</i> (S suffix)                |
|  | Puelche         | 1SG <i>ni-</i> (S prefix)                |
|  |                 | 1SG - <i>in</i> ~ - <i>an</i> (S suffix) |
|  |                 | 2SG <i>ki-ma-w</i> (personal pronoun)    |
|  |                 | 2SG <i>mu-</i> ~ <i>mi-</i> (S prefix)   |
|  |                 | 2SG - <i>ki-ma</i> (possessive suffix)   |

The distribution of most of these forms in the respective proto-languages or languages is of the type that, following Nichols & Peterson (2005: 546), we can call paradigmatic. An *n*-form and an *m*-form are opposed in a same form class, in the same number (singular), if a number distinction exists, and form a strict paradigm (e.g. *nīs-* : *mīs-*). This type of distribution, according to Nichols & Peterson (1996: 338), is strong evidence of a non-causal combination of the *n*- and *m*-forms, because it specifies particular consonants in particular positions of a multi-cell paradigm like the following. (The fillers of this scheme are Karok forms.)

|     | INDEPENDENT | POSSESSIVE                  |
|-----|-------------|-----------------------------|
| 1SG | <i>ná·</i>  | <i>nani-</i> ~ <i>nini-</i> |
| 2SG | <i>ʔi·m</i> | <i>mi-</i>                  |

Just Proto-Mixe-Zoquean, Puelche, and Takelma have a loose *n* : *m* paradigm straddling a combination of different form classes.

The presence of an *n* '1(SG) : *m* '2(SG)' (strict) paradigm in four further proto-languages cannot be excluded.

- (1) NORTH AMERICAN PACIFIC RIM: Proto-Utian.
- (2) MESOAMERICA & CENTRAL AMERICAN ISTHMUS: Proto-Chibchan.
- (3) CENTRAL & NORTHERN SECTORS OF WESTERN SOUTH AMERICA: Proto-Timotean, Proto-Uru-Chipaya.

### 5. Languages without *n-* or *m-*forms

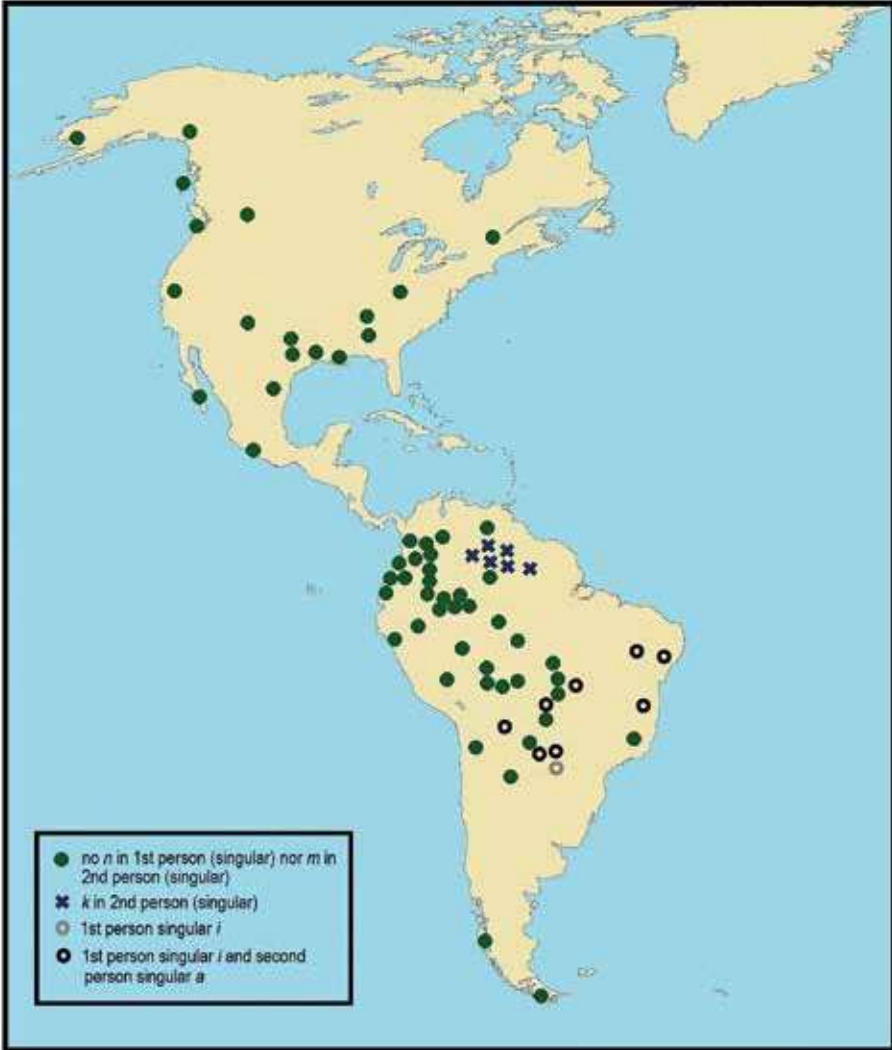
Map 6 shows the geographical localization of the 72 proto-languages and isolates of my sample that do not appear to have neither an *n*-form for first person (singular) nor an *m*-form for second person (singular).

- (1) NORTH AMERICA: Atakapan, Chitimacha, Cuitlatec, Haida, Kutenai, Proto-Athabaskan-Eyak-Tlingit, Proto-Caddoan, Proto-Chimakuan, Proto-Comecrudan, Proto-Eskimo-Aleut, Proto-Iroquoian, Proto-Muskogean, Proto-Siouan-Catawba, Tonkawa, Waikuri, Yana, Yuchi, Zuni.
- (2) SOUTH AMERICA: Andaquí, Atacameño, Aʔiwa, Betoj-Jirara, Camsá, Canichana, Cayubaba, Esmeraldeño, Fulniô, Guaicuruan, Guamo, Guató, Hodi, Irantxe, Itonama, Kwaza, Lule, Máku, Matanawí, Mochica, Movima, Proto-Nuclear Macro-Je, Páez, Payaguá, Pirahã, Proto-Arawan, Proto-Barbacoan, Proto-Boran, Proto-Bororoan, Proto-Cahuapanan, Proto-Chiquitano-Sansimoniano, Proto-Chocoan, Proto-Harakmbut, Proto-Huitotoan, Proto-Kamakanan, Proto-Karirian, Proto-Katukinan, Proto-Mascoyan, Proto-Matacoan, Proto-Peba-Yaguan, Proto-Qawasqaran, Proto-Sáliban, Proto-Tikuna-Yurí, Proto-Tiniguan, Proto-Yanomami, Proto-Zaparoan, Puri-Coroado, Sapé, Taruma, Taushiro, Trumai, Urutani, Yámana, Yurumanguí.

The map also highlights the distribution, in two distinct areas of South America, of three competing pronominal forms: a *k*-form for second person or second person singular, an *i*-form (or *j*) for first person singular, and *a*-form (or mid front *ɛ* or *e*, in a few cases) for second person singular. The last two forms constitute a pronominal mini-set in various languages (cf. Greenberg 1987: 44-46).

Second-person *k* is found in some languages of the western sector of the Guianas: the three languages that compose the Sáliban family and five isolates (Hodi, Máku, Sapé, Taruma, and Urutani; see section 3 of Appendix B).

The *i* : *a* mini-set is widespread in the central and southern sectors of the Brazilian Highlands and in the Gran Chaco. It is attested in eight independent lineages (Bororoan, Chiquitano-Sansimoniano,



**Map 6.** No *n* in first person (singular) nor *m* in second person (singular) (Americas).

Fulniô (isolate), Payaguá (isolate), Kamakanan, Karirian, Matacoan, and Nuclear Macro-Je). The component *i* of this set can also be observed in the isolate Guachí of the Paraguay River area, in Proto-Guaicuruan of Gran Chaco (see section 4 of Appendix B), in Kanoê, an isolate spoken in southern Rondônia that has an *m*-form for second person singular (see section 2 of Appendix B), and in Proto-Zamucoan, also in Gran Chaco, that, interestingly, has both an *m*-form and an

*a*-form for second person singular (see again section 2 of Appendix B).<sup>10</sup>

It is unclear how the South American *i : a* pronominal system should be interpreted historically, but its well-defined geographical distribution is hardly an accident.

Campbell (1994: 8) points out that Greenberg's claim about the *n : m* pattern as diagnostic for Amerind as a whole is not helped not only by those "Amerind" languages which have neither *n* nor *m* in first-person or second-person pronoun forms, but also by those "Amerind" languages whose behavior is the reverse of expectations, with second-person *n* and/or first-person *m*. In the rest of this section, we will try to evaluate in which proportions the *n*-forms for second person or second person singular and the *m*-forms for first person or first person singular are widespread in Native America through the proto-languages and languages of my sample.

The *n*-forms for '2(SG)' I noted, listed in section 5 of Appendix B, occur in the proto-languages of nine families and in fifteen isolates (24 families).

- (1) NORTH AMERICA: Kutenai, Molala, Proto-Atakapan, Proto-Eskimo-Aleut, Proto-Totonacan, Proto-Tsimshianic, Siuslaw, Tonkawa, Yana, Yuchi.
- (2) SOUTH AMERICA: Canichana, Kwaza, Leco, Máku, Movima, Mure, Proto-Barbacoan, Proto-Cahuapanan, Proto-Guahiboan, Proto-Harakmbut, Proto-Tupian, Taruma, Yaruro, Yurumangúí.

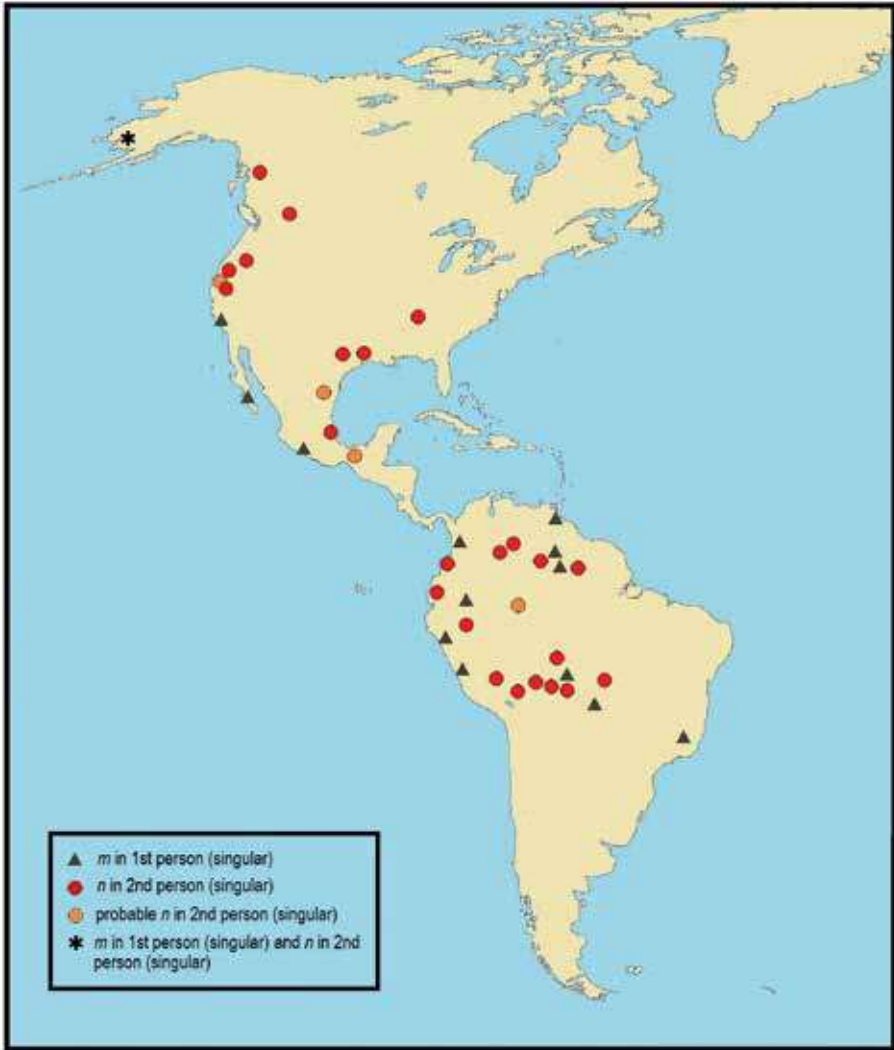
The attribution of an *n*-form for '2(SG)' to three proto-languages of North America (Proto-Comecrudan, Proto-Coosan, and Proto-Mixe-Zoquean) and one proto-language of South America (Proto-Katukinan) is uncertain (see again section 5 of Appendix B).

The *m*-forms for '1(SG)' I observed in my sample, listed in section 6 of Appendix B, occur far less than the *n*-forms for '2(SG)'. They appear in the proto-languages of five families and in nine isolates (fourteen families).

- (1) NORTH AMERICA: Cuitlateco, Proto-Chumashan, Proto-Eskimo-Aleut, Waikuri.
- (2) SOUTH AMERICA: Itonama, Mochica, Proto-Bororoan, Proto-Chocoan, Proto-Quechuan, Puri-Coroado, Sapé, Urutani, Waorani, Warao.

The geographical distribution of the *n*-forms for second person (singular) and *m*-forms for first person (singular) is shown in Map 7.





**Map 7.** *n* in second person (singular) (Americas) and/or *m* in first person (singular)

Curiously, the proto-languages and isolates that have *n* ‘2(SG)’ or *m* ‘1(SG)’ in North America are predominantly localized in coastal areas, both of the Pacific and the Atlantic, while in South America all are concentrated in the northern half of the continent. Only in Proto-Eskimo-Aleut we find a (non paradigmatic) combination of *m* ‘1(SG)’ and *n* ‘2(SG)’.<sup>11</sup> This removes the possibility of attributing some his-

torical significance to their combination. The presence in my sample of *n*-forms for ‘2(SG)’ and *m*-forms for ‘1(SG)’, besides *n*-forms for ‘1(SG)’ and the *m*-forms ‘2(SG)’, rather confirms Campbell’s (1994: 4) observation that personal pronouns, like other basic elements of languages, tend to contain nasals and that this tendency makes itself felt in both first person and second person.<sup>12</sup> Finally, the fact that a little under half (44%) of proto-languages and isolates of my sample have neither *n* nor *m* in first or second (singular) pronominal forms indicates that the claim for ubiquity of first-person *n* and second-person *m* in the languages of Native America (*pace* Greenberg 1987: 49) is largely overstated.

## 6. *n*- and *m*-forms in Eurasia

Based on their sample of 173 languages of the world, Nichols & Peterson (1996: 361) indicate that, although nasals in general are of high frequency in pronominal forms (see below section 8), the *n* : *m* pronominal system is exceedingly rare outside western America, being appreciably widespread only in northern coastal New Guinea.

In this section, I propose a survey of the geographical distribution of first-person *n* and second-person *m* in Eurasia based on a sample that, like the American one, where possible, contains one representative from each generally accepted known family of the region that, in the case of a multi-member family, coincides with its proto-language.

Given that, at the current state of knowledge, the human colonization of the Americas emanated from populations (or, perhaps, just one group; cf. Goebel, Waters & O’Rourke (2008)) originating in Siberia, it appeared to me opportune to extend research of *n* : *m* pronominal systems to the entire Eurasian landmass in search of possible extensions of the (putative) American *n* : *m* paradigm. Keeping in mind that the languages of Eurasia typically have a number distinction in their pronominal forms, I took into account only first-person singular and second-person singular pronominal forms for this survey.

The families of Eurasia that were examined — 35 in all including 14 isolates — are those recognized by *Glottolog* 3.0.<sup>13</sup> Their list and localization are supplied by Table 5 in Appendix A. Of the proto-languages of most multi-member families, we today can avail of more or less wide reconstructions.<sup>14</sup> The living language isolates, except Hruso and Shompeng, have been, at least, adequately described both grammatically and lexically. Of Shompeng only elementary grammatical

data and few words were gathered. Among the recorded words there is the first-person singular personal pronoun, but not the second-person singular one. Of the extinct isolates relatively little is known, but of these, at least, we know the first and second-person singular personal pronouns.

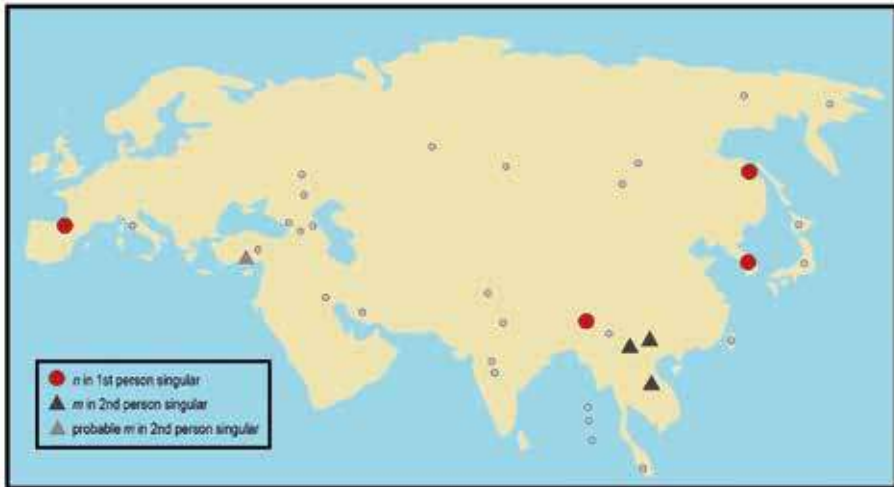
Based on my sample, only four language isolates in Eurasia have at least one *n*-form for first person singular: Basque, Hruso, Korean, and Nivkh (see section 7 of Appendix B). And just for three proto-languages of Eurasia (the Southeast Asian mainland to be precise) an *m*-form for second person singular is reconstructable: Proto-Austroasiatic, Proto-Hmong-Mien, and Tai-Kadai. For a further proto-language, Proto-Hurro-Urartian, the presence of a second-person singular *m*-form is probable, but not certain (see section 8 of Appendix B). No proto-language or known isolate of Eurasia has therefore a (loose or strict) pronominal paradigm with first-person singular *n* and second-person singular *m*. This is definitively not a Eurasian pronominal paradigm.

Map 8 shows the location of the proto-languages and isolates of my Eurasian sample. Also in this map the proto-languages of families with a wide territorial extension are placed near their supposed Urheimat (see Table 6 in Appendix A).

Map 9 shows the geographical distribution of the few Eurasian *n*- and *m*-forms for first person singular and second person singular in the sample.



**Map 8.** Location of the proto-languages of the linguistic families and of the language isolates of Eurasia. (Proto-languages and language isolates are identified by the same numbers in Table 5 in Appendix A.)



**Map 9.** *n* in first person singular and *m* in second person singular (Eurasia).

### 7. Frequency of the *n*- and *m*-forms in the Americas

Tables 9 and 10 give, per area, the frequencies with which first-person (singular) *n*, second-person (singular) *m*, first-person (singular) *m*, and second-person (singular) *n* occur in my sample of American proto-languages and languages. The areas taken into account are the seven areas of the New World in which language isolates are or were once spoken and/or the Urheimat of at least one language family (supposedly) lies (section 3): 1. North American Pacific Rim; 2. Great Plains, Southern Plains, Southeast United States, Midwest United States, Saint Lawrence Lowland, and Newfoundland; 3. Mesoamerica and Central American Isthmus; 4. Central and northern sectors of western South America; 5. São Francisco basin and eastern and southeast Atlantic regions of Brazil; 6. Eastern Pampas; 7. Southern Andes and southern Patagonia. All percentages given are rounded.

Given the relatively low numbers of my sample and due to the necessity of presenting data that may be immediately compared with the data supplied by Nichols and Peterson's 1996 article, no more elaborate statistical analysis of the American *n*- and *m*-forms will be attempted. The simple comparison of the percentages that follows with those supplied by Nichols & Peterson (1996) allows the reader to gather the essence of the matter.

*First-person n and second-person m in Native America: a fresh look*

| AREA                       | NO. OF PROTO-LANGUAGES<br>AND ISOLATES | <i>n</i> |         | <i>m</i> |     |
|----------------------------|--|----------|---------|----------|-----|
|                            |  | No.      | %       | No.      | %   |
| 1                          | 36                                     | 18-21    | 50%-58% | 3        | 8%  |
| 2                          | 21                                     | 8-9      | 38%-43% | 0        | 0%  |
| 3                          | 13                                     | 7-9      | 54%-69% | 1        | 8%  |
| 4                          | 89                                     | 14-17    | 16%-19% | 9        | 10% |
| 5                          | 5                                      | 0        | 0%      | 1        | 20% |
| 6                          | 1                                      | 0        | 0%      | 0        | 0%  |
| 7                          | 7                                      | 3        | 43%     | 0        | 0%  |
| NORTH AMERICA <sup>a</sup> | 70                                     | 33-39    | 47%-56% | 4        | 6%  |
| SOUTH AMERICA <sup>b</sup> | 102                                    | 17-20    | 17%-20% | 10       | 10% |
| TOTAL                      | 172                                    | 50-59    | 29%-34% | 14       | 8%  |

Notes. <sup>a</sup> Areas 1-3. <sup>b</sup> Areas 4-7.

**Table 9.** Frequencies of *n* and *m* in pronominal forms for '1(SG)' in the proto-languages of families and in the language isolates of the Americas.

| AREA          | NO. OF PROTO-LANGUAGES<br>AND ISOLATES | <i>n</i> |         | <i>m</i> |         |
|---------------|--|----------|---------|----------|---------|
|               |  | No.      | %       | No.      | %       |
| 1             | 36                                     | 6-7      | 17%-19% | 21-22    | 58%-61% |
| 2             | 21                                     | 3-4      | 14%-19% | 4        | 19%     |
| 3             | 13                                     | 1-2      | 8%-15%  | 6-7      | 46%-54% |
| 4             | 89                                     | 14-15    | 16%-17% | 23-26    | 26%-29% |
| 5             | 5                                      | 0        | 0%      | 0        | 0%      |
| 6             | 1                                      | 0        | 0%      | 1        | 100%    |
| 7             | 7                                      | 0        | 0%      | 3        | 43%     |
| NORTH AMERICA | 70                                     | 10-13    | 14%-19% | 31-33    | 44%-47% |
| SOUTH AMERICA | 102                                    | 14-15    | 14%-15% | 27-30    | 26%-29% |
| TOTAL         | 172                                    | 24-28    | 14%-16% | 58-63    | 34%-37% |

**Table 10.** Frequencies of *n* and *m* in pronominal forms for '2(SG)' in the proto-languages of families and in the language isolates of the Americas.

The two tables reaffirm what we observed in section 3. The presence of *n*-forms for first person (singular) is very strong in North America (47%-56%), but not particularly significant in South America (17%-20%); the presence of *m*-forms for second person (singular) is very high in the Pacific Rim region of North America (58%-61%), weaker in the rest of the continent (29%-32% on the whole), and even weaker in South America (26%-29%).

Tables 11 and 12 below show the frequencies of first-person and second-person singular personal pronouns with *n* or *m* in initial position observed by Nichols & Peterson (1996: 351) in their sample of 173 languages covering the world.

| AREA                       | NO. OF LANGUAGES | <i>n</i> |     | <i>m</i> |     |
|----------------------------|------------------|----------|-----|----------|-----|
|                            |                  | No.      | %   | No.      | %   |
| W North America            | 29               | 13       | 45% | 0        | 0%  |
| E North America            | 13               | 2        | 15% | 1        | 8%  |
| Mesoamerica                | 10               | 1        | 10% | 0        | 0%  |
| South America              | 19               | 3        | 16% | 0        | 0%  |
| Africa                     | 20               | 5        | 25% | 5        | 25% |
| Europe                     | 12               | 1        | 8%  | 7        | 58% |
| Northern Asia              | 15               | 2        | 13% | 5        | 33% |
| S & SE Asia                | 8                | 0        | 0%  | 0        | 0%  |
| New Guinea                 | 28               | 9        | 32% | 1        | 4%  |
| Australia                  | 19               | 0        | 0%  | 0        | 0%  |
| NORTH AMERICA <sup>a</sup> | 52               | 16       | 31% | 1        | 2%  |
| AMERICAS <sup>b</sup>      | 71               | 19       | 27% | 1        | 1%  |

Notes. <sup>a</sup> Including Mesoamerica. <sup>b</sup> W North America, E North America, Mesoamerica, and South America.

**Table 11.** Frequencies of initial *n* and *m* in independent personal pronouns for ‘1sg’ in Nichols & Peterson’s (1996) sample of 173 languages of the world.

| AREA            | NO. OF LANGUAGES | <i>n</i> |     | <i>m</i> |     |
|-----------------|------------------|----------|-----|----------|-----|
|                 |                  | No.      | %   | No.      | %   |
| W North America | 29               | 1        | 3%  | 14       | 48% |
| E North America | 13               | 3        | 23% | 1        | 8%  |
| Mesoamerica     | 10               | 0        | 0%  | 2        | 20% |
| South America   | 19               | 1        | 5%  | 4        | 21% |
| Africa          | 20               | 2        | 10% | 3        | 15% |
| Europe          | 12               | 0        | 0%  | 0        | 0%  |
| Northern Asia   | 15               | 1        | 7%  | 0        | 0%  |
| S & SE Asia     | 8                | 3        | 37% | 1        | 12% |
| New Guinea      | 28               | 8        | 29% | 3        | 11% |
| Australia       | 19               | 4        | 21% | 0        | 0%  |
| NORTH AMERICA   | 52               | 4        | 8%  | 17       | 33% |
| AMERICAS        | 71               | 5        | 7%  | 21       | 30% |

**Table 12.** Frequencies of initial *n* and *m* in independent personal pronouns for ‘2sg’ in Nichols & Peterson’s (1996) sample of 173 languages of the world.

Table 11 shows a high frequency of first-person singular pronouns with initial *n* only for the western region of North America (45%), Eastern North America and Mesoamerica (including the Central American Isthmus) having a frequency of these forms at least a third lower (respectively 15% and 10%). Table 12, in accordance with my data, indicates that second-person singular personal pronouns with initial *m* also predominate in western North America (48%) and do not make a particularly strong showing in the rest of the Americas where they appear with a frequency (17% on the whole) which is not too far from the frequencies they have in other areas of the world (Africa, South and Southeast Asia, and New Guinea).

In Table 13, I reported the frequencies that the *n* : *m* paradigm (whether strict or loose) has in my sample both with the values '1(SG)' for *n* and '2(SG)' for *m* and the values '2(SG)' for *n* and '1(SG)' for *m*.

| AREA          | NO. OF PROTO-LANGUAGES<br>AND ISOLATES | <i>n</i> '1(SG)' : <i>m</i> '2(SG)' |         | <i>n</i> '2(SG)' : <i>m</i> '1(SG)' |    |
|---------------|--|-------------------------------------|---------|-------------------------------------|----|
|               |  | No.                                 | %       | No.                                 | %  |
| 1             | 36                                     | 13-14                               | 36%-39% | 1                                   | 3% |
| 2             | 21                                     | 4                                   | 19%     | 0                                   | 0% |
| 3             | 13                                     | 3-4                                 | 23%-30% | 0                                   | 0% |
| 4             | 89                                     | 2-4                                 | 2%-4%   | 0                                   | 0% |
| 5             | 5                                      | 0                                   | 0%      | 0                                   | 0% |
| 6             | 1                                      | 0                                   | 0%      | 0                                   | 0% |
| 7             | 7                                      | 2                                   | 29%     | 0                                   | 0% |
| NORTH AMERICA | 70                                     | 20-22                               | 29%-31% | 1                                   | 1% |
| SOUTH AMERICA | 102                                    | 4-6                                 | 4%-6%   | 0                                   | 0% |
| TOTAL         | 172                                    | 24-28                               | 14%-16% | 1                                   | 1% |

**Table 13.** Frequencies of the *n* : *m* paradigm with the values '1(SG)' and '2(SG)' and '2(SG)' and '1(SG)' in the proto-languages of families and in the language isolates of the Americas.

The *n* '1(SG)' : *m* '2(SG)' paradigm is found in over one third (36%-39%) of the proto-languages and isolates of area 1, the North American Pacific Rim region. Its occurrence in about one fifth (19%) of the proto-languages and isolates of area 2, including the Great Plains, the Southern Plains, the Southeast United States, the Midwest United States, the Saint Lawrence Lowland, and Newfoundland, is also noteworthy as is the even greater occurrence which can be observed in area 3, Mesoamerica and Central American Isthmus (23%-30%). In area 4, covering the central and

northern sectors of western South America, where about half (52%) of the known American families are located (precisely 89 out of 172, as Table 13 indicates), the paradigm has a decisively low frequency of occurrence: 4% at most. In the São Francisco basin and in the Eastern and Southeast Atlantic regions of Brazil, the paradigm is absolutely absent, while its occurrence in two sole isolates of the Southern Andes-Southern Patagonia area (Mapuche and Puelche) determines that its frequency in this area is equal to 29%.

The  $n$  '2(SG) :  $m$  '1(SG)' paradigm has an insignificant presence in the Pacific Rim region of North America (where it is found only in one of its 36 commonly recognized families: Proto-Eskimo-Aleut) and is absolutely absent in the rest of the New World.

The frequency of occurrence of the  $n$  '1(SG) :  $m$  '2(SG)' paradigm in all of North America given above (39%) is rather close to that of the  $n$  '1SG' :  $m$  '2SG' paradigm in the continent based on Nichols & Peterson's (1996) sample of 71 modern American languages (35%; see Table 14). The frequency of  $n$  :  $m$  in the North American Pacific Rim region computed in Table 13 (39%) is, in particular, almost identical to what Nichols & Peterson (1996: 357-358) indicate for Western North America (38%). But, if we pass from North America to South America, we may note a profound difference between the frequency of the  $n$  :  $m$  paradigm established on the basis of my sample with that indicated by Nichols and Peterson: 4%-6% vs. 32%.

| AREA                       | NO. OF LANGUAGES | NO. | %   |
|----------------------------|------------------|-----|-----|
| W North America            | 29               | 11  | 38% |
| E North America            | 13               | 1   | 8%  |
| Mesoamerica                | 10               | 6   | 60% |
| South America <sup>a</sup> | 19               | 6   | 32% |
| Africa                     | 20               | 2   | 10% |
| Europe                     | 12               | 0   | 0%  |
| Northern Asia              | 15               | 0   | 0%  |
| S & SE Asia                | 8                | 1   | 12% |
| New Guinea                 | 28               | 4   | 14% |
| Australia                  | 19               | 0   | 0%  |
| <br>                       |                  |     |     |
| NORTH AMERICA              | 52               | 18  | 35% |
| AMERICAS                   | 71               | 24  | 34% |

*Note.* <sup>a</sup> South America also includes the Central American Isthmus and, specifically, the Guaymí (Ngäbere) language of Panama (Chibchan family).

**Table 14.** Frequencies of the  $n$  '1SG' :  $m$  '2SG' paradigm in Nichols & Peterson's (1996) sample of 173 languages of the world.



Besides, if we observe in detail the percentages concerning the southernmost zone of the North American continent supplied by Tables 13 and 14, we may also notice that the frequency of the *n* '1SG' : *m* '2SG' paradigm in Mesoamerica indicated by Nichols and Peterson's sample is at least twice the frequency that I indicated for the *n* '1(SG)' : *m* '2(SG)' paradigm in area 1 (Mesoamerica and Central American Isthmus): 60% vs. 23%-30%. These discrepancies with my data are also imputable both to the scarce representativeness of the American sample surveyed by Nichols and Peterson (six languages for Mesoamerica and only six languages of the entire South America) and to the two authors' choice not to examine the diachronic antecedents of the *n*- and *m*-forms found. This choice also allows for six secondary *n* : *m* paradigms (Coatlán Mixe and Highland Chontal in Mesoamerica and Axininca Campa, Cashinahua, Hixkaryana, and Mískito in South America; see note 6) to be counted.<sup>15</sup>

Campbell (1997b: 246-247) refers that, in a worldwide sample of 333 languages, Matthew Dryer found that 7% of the non-Amerindian languages — i.e. 17 out of the remaining 252 languages — had both an *n* in first person and an *m* in second person, either with both as singular or both as plural. In Dryer's sample, 17% of the languages from Greenberg's Amerind (14 out of 81 languages) had this paradigm. North America, based on my sample, is considerably above these average figures (30%-31%), while South America is below both average figures (4%-6%).

## 8. Conclusions

Pronouns are among the most stable elements of basic vocabulary (Dixon 1997: 22; Thomason & Kaufman 1988: 75, 76; Muysken 2008: 95; Parkvall 2008: 238; Matras 2009: 203-208; Heine & Song 2010: 117), at least in several language families. Pronominal forms can be reconstructed for proto-languages of deep-time language families like Indo-European (Katz 1998), Afroasiatic (Ehret 1995, Satzinger 2004), Niger-Congo (Babaev 2013), and Austronesian (Blust 1977, Ross 1996), and this means that we have evidence of a remarkable time-stability of pronouns over a period (apparently, in the case of Niger-Congo) of 10,000-12,000 years. This does not imply, however, that all similarities that can be observed among the pronominal forms of the languages of the world are always and necessarily due to genealogical relationships. As indicated by Campbell (1994: 3-4, 1997a: 344-346), Goddard & Campbell (1994: 196-197), Willerman

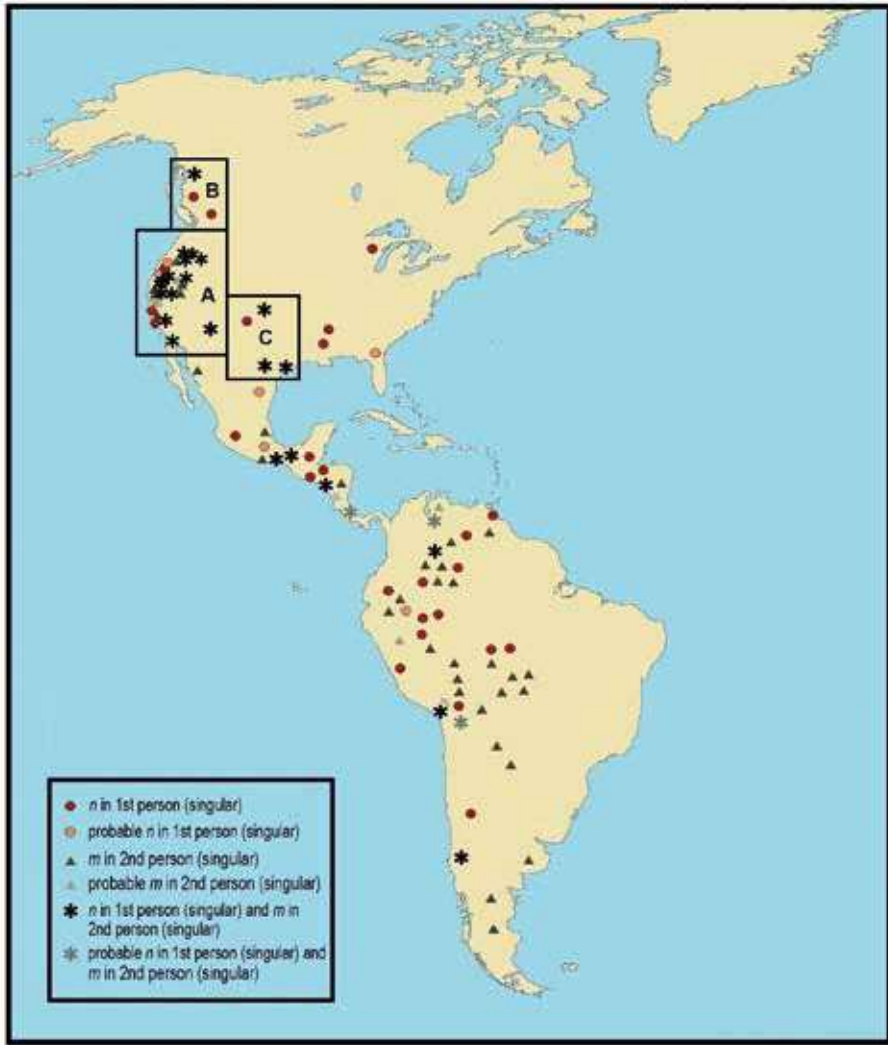
(1994: 38), Gordon (1995), Sasse (2015: 197), and other authors, most languages of the world construct their pronominal systems around a small set of basic consonants (including *n*, *m*, *t*, *k*, and *s*), often with the same functions across languages. Ergo, we cannot always establish with certainty whether such similarities result from historical continuity or from the fact that pronouns, like other closed-class forms (e.g. adpositions, conjunctions, and spatial adverbs), show a permanent tendency to simplify their structure and to become restricted to the most basic phonemes (and syllable types).

Having said this, we can neither ignore that the frequency with which the *n* '1(SG)' : *m* '2(SG)' paradigm occurs in the Native languages of North America, specifically those of the Pacific Rim area of the continent, is such that leads us to exclude that it is simply due to distributional universal tendencies or random chance. According to the sample used in this article, including representatives from each (proven) American family (usually the proto-language in the case of a multi-member family), an *n* : *m* pronominal paradigm belongs to the 36%-39% of the independent linguistic lineages of the North American Pacific Rim. And this value is more than five times greater than that 7% of non-American languages that have the *n* : *m* paradigm, either with both as singular or both as plural, found by Dryer in his sample of 333 languages of the world (section 7). The density of the *n* : *m* paradigm is also definitely high in the area of North America that includes the Great Plains, the Southern Plains, the Southeast United States, the Midwest United States, the Saint Lawrence Lowland, and Newfoundland (19%) and in the Mesoamerica-Central American Isthmus area (23%-30%), but it is definitely low in South America (4%-6%). These data therefore indicate that, rather than a "western" or "Pacific Rim" feature, as claimed by Nichols & Peterson (1996, 1998, 2005), the *n* : *m* paradigm is a North American feature not restricted to its western portion. Whether this pronoun system forms a single large areal cluster is, however, unclear. The sample I used reveals that the greatest density of *n*- and *m*-forms is in the area formed by the southern half of the Northwest Coast and the adjacent California (including northern Baja California) (section 3), but the occurrence of *n*- and *m*-forms in Proto-Uto-Aztecan — whose Urheimat is usually placed in the territory between present-day Southeastern California, Arizona, and the northwest part of Sonora (Gross 1968: 17; Fowler 1983: 228-234; Foster 1996: 91, 93; Campbell 1997: 137)<sup>16</sup> — permits us to identify a wider uninterrupted zone of diffusion of the *n* : *m* paradigm that, in its southern part, (probably) extends beyond the Colorado River (the area A of Map 10).

The presence of *n*- and *m*-forms for first person singular and second person singular in Proto-Tsimshianic suggests, in addition, that the area where the pronominal paradigm is spread may also extend to the northern half of the Northwest Coast until the Nass River, encompassing the Urheimaten of the Wakashan and Salishan families (whose proto-languages, perhaps not by chance, have an *n*-form for first person singular; see the area B of Map 10), while the presence of *n* and *m* in Proto-Kiowa-Tanoan, whose linguistic area is placed by Ortman (2012: 158) in an upland environment “most likely adjacent to the Rocky Mountains”, likely in southern Colorado according to Shaul (2014: 105), and in the extinct Coahuilteco and Karankawa, once spoken, both, in the southern sector of the Southern Plains, would seem to indicate an extension of the area under discussion as far as this region of North America (area C of Map 10). Supposing that the *n*-forms occurring in Tunica (Louisiana and Mississippi; *-ni* ‘1’, *-ka’ni* ‘id.’, and *-ʔan* ‘id.’), Natchez (also in Louisiana and Mississippi; *-niš* ‘1’ and *-n ~ -ni* ‘id.’), and Timucua (Georgia and Florida; *ni-* ‘1’ and *-na* ‘id.’), which lack second-person singular *m*, and the various *m*-forms of Seri (Sonora; *me* ‘2’, *mi-* ~ *ma-* ‘id.’, *ma* ‘2sg’, *m-* ‘id.’, etc.), which lack first-person singular *n* (sections 2.1 and 2.2), are the residue of an ancient *n : m* set of which, with time, a component was replaced or underwent phonological reshaping,<sup>17</sup> we might also assume that the area of diffusion of the *n : m* paradigm may continuously extend in a southerly direction until the central sector of the Sonoran desert and in an easterly direction in the Southeastern United States. Staying within the sphere of pure hypotheses, we might also suppose that the *n : m* paradigm is an even wider areal cluster that extends until Mesoamerica or, perhaps, even beyond, until the Central American Isthmus (Proto-Chibchan). In order to suppose this we must, however, ignore that, in light of current knowledge, it is impossible to demonstrate that the *n : m* paradigm had in the past an uninterrupted distribution from California and/or the Southern Plains to Mesoamerica.

Surely Nichols and Peterson are right in concluding in their 1996 article that the *n : m* pronominal paradigm is not an evidence of the relatedness of American languages given its uneven geographical distribution in the New World.

“The *n : m* pronominal paradigm is widely regarded as the strongest single piece of evidence for Amerind as a genetic grouping, but in fact the geographical distribution of this paradigm, or for that matter of its component second person *m*, calls Amerind into question



**Map 10.** Three areas of diffusion of the  $n : m$  pronominal paradigm in North America.

and may be sufficient to refute it. If Amerind were a genetic reality and the  $n : m$  paradigm a marker of it, then the marker should have a fairly even distribution over all of Amerind and should be found only there. We have shown that the  $n : m$  paradigm is not evenly distributed over Amerind, or over the New World more generally; it is a western American phenomenon, and the difference in its frequency in western and eastern North America, or western and eastern

America more generally, is statistically highly significant.” (Nichols & Peterson 1996: 367)

But in light of the data presented in this article, as indicated above, it does not seem legitimate to consider the *n* : *m* paradigm a western American phenomenon, as they indicate. The paradigm, observed through the proto-languages of all multi-member families and all isolates of the New World we know, appears essentially as a North American phenomenon that cannot be circumscribed precisely to a single area and is deeply-rooted only in the southern half of the Northwest Coast and in the neighboring California (Map 10). Not being possible to establish whether the conforming families of North America constitute a single areal cluster or not, it is impossible to draw a unique and sure conclusion about the spreading of the *n* : *m* paradigm in the continent. Limited to the southern half of the Northwest Coast and California, where the density of *n* : *m* is strikingly high, and, very likely, the close Southwest (to the presumed Uto-Aztecan Urheimat), we may suppose that the pronominal paradigm is a clear indicator of prehistoric contacts. But the question of whether these contacts have the form of a hypothetical proto-language or of a convergence between (perhaps few or very few) formerly distinct language branches is not answerable based on just the present-day knowledge of the history of the linguistic families of these areas. With a lesser degree of certainty, we can also speculate that such contacts also regarded languages once spoken in the northern half of the Northwest Coast (cf. Proto-Tsimshianic *\*=nə* ‘1SG’ : *\*=mə* ‘2SG’) and in the Southern Plains (cf. Coahuilteco *niw-* ‘1SG(>3O)’ : *may-* ~ *mi-* ‘2SG(>3O)’, Karankawa *na-* ‘1SG’ : *m-* ‘2SG’, and Proto-Kiowa-Tanoan *\*nq* ‘I’ : *\*wim* ‘you’ (sg.)), but, as far as Mesoamerica, the Central American Isthmus, and the whole South American region are concerned, there are not strong elements which prevent us from thinking that the *n* : *m* paradigm we find in some proto-languages and isolates of these areas (four or, at most, six in all of South America) are the result of parallel but unconnected developments.

*Abbreviations and conventions*

|       |  |         |          |
|-------|--|---------|----------|
| 1     | first person                               | pl.     | plural   |
| 2     | second person                              | S       | subject  |
| 3     | third person                               | sg, sg. | singular |
| O     | object                                     |         |          |
| X > Y | subject/agent X acting on object/patient Y |         |          |

Notes

<sup>1</sup> The widespread *n* of first person in various American Indian languages was well-known by the 1870s (Sayce 1874-1875: 216, Tolmie & Dawson 1884: 128-129, Brinton 1888: 215-216).

<sup>2</sup> “L’elenco è riuscito tutt’altro che completo per insufficienza del materiale a nostra disposizione, ma basta certamente per dare un’idea della grande diffusione di questi elementi antichissimi ed essenziali”.

<sup>3</sup> “Dalle regioni più settentrionali dell’America (...) fino all’estremità meridionale del nuovo continente, alla Terra del Fuoco” (...) (Trombetti 1905: 208) [from the most northern regions of America (...) to the southern extreme of the new continent, to the Tierra del Fuego].

<sup>4</sup> Also published online, in 2013, at <http://wals.info/chapter/136>.

<sup>5</sup> The proto-languages of the families to which the following languages belong do not seem to have a complete *n* : *m* paradigm: Axininca Campa\* (Arawakan), Cashinahua\* (Panoan), Coatlán Mixe\* (Mixe-Zoquean), Highland Chontal\* (Tequistlatecan), Mískito\* (Misumalpan), Wari\* (Chapakuran), Wichi/Mataco\* (Matacoan), Xokleng (Nuclear Macro-Je), Yurok\* (Algic). (The asterisk signals languages that were also included in Nichols and Peterson’s 1996 sample.) In addition, Rama could also have a secondary *n* : *m* system given that the attribution of the *n* : *m* configuration to Proto-Chibchan is uncertain (see sections 1 and 2 of Appendix B).

<sup>6</sup> Various American languages make no number distinctions in pronominal forms.

<sup>7</sup> As this table indicates, the location of an Urheimat is often uncertain or particularly generic. Considerable confidence regarding the location of the Urheimat of a language family comes from multiple lines of linguistic, genetic, and archaeological evidence, even when the precise contours of a proto-language are not firmly established. For most of the American families, however, this kind of multidisciplinary evidence is currently not available.

<sup>8</sup> The attribution of an *m*-form for second person to Proto-Algic is particularly uncertain. I have not taken account of it in the rest of this article. Proulx (1985: 86) reconstructs for Proto-Algic a second-person actor suffix \*-Vm(?) based on Yurok and Wiyot forms, but he advises that “[t]he Proto-Algic status of this reconstruction is in doubt, since *m* often signals the second person in Western languages [of North America], and there is no Algonquian cognate. The suffix may be borrowed” (*ibid.*). Note that Proulx (*ibid.*) also reconstructs for Proto-Algic a more plausible second-person actor suffix \*-Vt based on Proto-Algonquian \*-at ‘thou-her’ (conjunct) and Wiyot *-ət* ~ *-it* ~ *-at* ‘id.’ (indicative).

<sup>9</sup> The personal pronoun *ma* ‘you’ (sg.) contributes to a Tunica *n* : *m* set for Nichols & Peterson (1996: 357). This form, as the two authors indicate (p. 361), consists however in a bare stem \*(-)ma found in most personal pronouns of Tunica (cf. Haas 1941: 38).

<sup>10</sup> The *i*-form of Kanoê is the verbal subject marker *i*- ‘1’ (Bacelar 2004: 143). For Proto-Zamucoan, Ciucci & Bertinetto (2017: 315-317) reconstruct \*j- ‘1sg’, possessive prefix and first-person irrealis subject prefix, and \*a- ‘2sg’, with identical uses.

<sup>11</sup> Two secondary *m* ‘1sg’ : *n* ‘2sg’ sets can be found in Proto-Dakotan and its descendant languages: \*mj= : \*ni= (patient), from Proto-Siouan \*wĵ : \*yi- ~ \*yj, and \*mą : \*ni (possessor), from Proto-Siouan \*wą (?) : \*yi- ~ \*yj (Rankin et al. 2015).

<sup>12</sup> Note also that in some of the proto-languages and isolates of my sample we find first-person (singular) *n* alongside first-person (singular) *m* (Molala, Mure, Proto-Chumashan, Proto-Guahiboan, Proto-Quechuan, Warao, and, perhaps,

Proto-Mixe-Zoquean) or second-person (singular) *n* alongside second-person (singular) *m* (Proto-Totonacan, Proto-Tsimshianic, and Yaruro). Also note in Leco the presence of a second-person singular marker that has both an allomorph *-m* and an allomorph *-n* (sections 2 and 5 of Appendix B).

<sup>13</sup> The families of the sample do not include Afroasiatic because its Urheimat is usually located in Africa (North Africa, the Horn of Africa, or the Eastern Sahara). In vain, I looked for useful data in the (available descriptions of the) scarce and/or problematic material of various Eurasian *Restsprachen* of unknown or uncertain genealogical affiliation: Iberian, Pictish, Camunic, Rhaetic, North Picene, Cassite, Khitan, and others.

<sup>14</sup> Table 7, in Appendix A, supplies a list of the proto-languages in question and indicates for each of them one relevant work. There are no current detailed reconstructions of the proto-languages of the Dravidian, Japonic, Mongolic, Nakh-Daghestanian, and Tungusic families. We dispose, however, of etymological dictionaries and/or comparative studies of the phonologies of their modern descendants (see Table 8 in Appendix A). For the two languages of the Hurro-Urartian family, no etymological dictionary or in-depth comparative phonological study of has ever been prepared. A preliminary reconstruction of Proto-Great Andamanese forms is forthcoming (Blevins *forthcoming*).

<sup>15</sup> Although my computation of the *n*- and *m*-forms, in contrast with that of Nichols and Peterson's, does not assume strict morpheme-initial position of the nasal and also takes into account forms where the nasal is the first consonant but not initial, the frequencies of the *n* '1(SG) : *m* '2(SG)' paradigm in the whole of North America and in the whole of South America based on my sample (respectively 29%-31% and 4%-6%) are not superior to the frequencies of the *n* '1SG' : *m* '2SG' paradigm in the two continents based on Nichols and Peterson's sample.

<sup>16</sup> Shaul (2014: 184-217), however, presents an interesting broad argument for placing the homeland of Proto-Uto-Aztecan in the Central Valley of California and eastwards.

<sup>17</sup> This, hypothetically, could have happened with the second-person subject marker *pan-* ~ *pi-* ~ *pa-* of Natchez; cf. the case of Lower Pima *a:pi* 'you' (sg.) from Proto-Uto-Aztecan \**i(mi)* 'id.' mentioned in section 1.

### *Bibliographical references*

- Aikhenvald, Alexandra Y. 1999. The Arawak language family. In Dixon, Robert M. W. & Aikhenvald, Alexandra Y. (eds.), *The Amazonian languages*. Cambridge: Cambridge University Press. 65-106.
- Babev, Kirill V. 2009. Once again on the comparison of personal pronouns in proto-languages. *Journal of Language Relationship* 1. 37-48.
- Babev, Kirill V. 2013. *Nigero-kongolezskij prajazyk: ličnye mestoimenija*. Moscow: Jazyki Slavjanskoj Kultury.
- Bacelar, Laércio Nora 2004. *Gramática da língua kanoê*. PhD dissertation. Katholieke Universiteit Nijmegen.
- Blevins, Juliette *forthcoming*. Linguistic clues to Andamanese pre-history: understanding the North-South divide. In Güldemann, Tom; McConvell, Patrick & Rhodes, Richard (eds.), *Hunter-gatherers and linguistic history: a global perspective*. Cambridge: Cambridge University Press.
- Blust, Robert A. 1977. The Proto-Austronesian pronouns and

- Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics* 9(2). 1-15.
- Brinton, Daniel G. 1888. The language of palæolithic man. *Proceedings of the American Philosophical Society* 1888. 3-16.
- Campbell, Lyle 1994. Problems with the pronouns in proposals of remote relationships among Native American languages. In Langdon, Margaret (ed.), *Proceedings of the Meeting of the Society for the Study of the Indigenous Languages of the Americas and the Hoka-Penutian Workshop*. Berkeley: University of California. 1-20.
- Campbell, Lyle 1997a. Amerind personal pronouns: a second opinion. *Language* 73. 339-351.
- Campbell, Lyle 1997b. *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.
- Carvalho, Fernando O. de 2017. Fricative debuccalization and primary split in Terena (Arawak) historical phonology. *International Journal of American Linguistics* 83. 509-537.
- Chirikba, Viacheslav 1996. *Common West Caucasian: the reconstruction of its phonological system and parts of its lexicon and morphology*. Leiden: Research School CNWS.
- Cincius, Vera I. 1949. *Sravnitel'naja fonetika tunguso-man'čžurskich jazykov*. Leningrad: Učpedgiz.
- Cincius, Vera I. (ed.) 1975-1977. *Sravnitel'nyj slovar' tunguso-man'čžurskich jazykov: materialy k ètimologičeskomu slovarju*. Leningrad: Nauka. 2 vols.
- Ciucci, Luca & Bertinetto, Pier Marco 2017. Possessive inflection in Proto-Zamucoan: a reconstruction. *Diachronica* 34. 283-330.
- Décsey, Gyula 1998. *The Turkic protolanguage: a computational reconstruction*. Bloomington: Eurolingua.
- Dixon, Robert M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Dixon, Roland B. 1910. The Chimariko Indians and their language. *University of California Publications in American Archaeology and Ethnology* 5. 295-380.
- Ehret, Christopher 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian): vowels, tone, consonants, and vocabulary*. Berkeley, Los Angeles & London: University of California Press.
- Fowler, Catherine S. 1983. Some lexical clues to Uto-Aztecan prehistory. *International Journal of American Linguistics* 49. 224-257.
- Frellesvig, Bjarke & Whitman, John 2008. Evidence for seven vowels in proto-Japanese. In Frellesvig, Bjarke & Whitman, John (eds.), *Proto-Japanese: issues and prospects*. Amsterdam: John Benjamins. 15-41.
- Goddard, Ives & Campbell, Lyle 1994. The history and classification of American Indian languages: what are the implications for the peopling of the Americas. In Bonnicksen, Robson & Steele, D. Gentry (eds.), *Method and theory for investigating the peopling of the Americas*. Corvallis: Oregon State University. 189-207.
- Goebel, Ted; Waters, Michael R. & O'Rourke, Dennis H. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319. 1497-1502.
- Gordon, Matthew 1995. The phonological composition of personal pro-



- nouns: implications for genetic hypotheses. *Berkeley Linguistics Society* 21. 117-128.
- Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.
- Gross, James A. 1968. Culture-historical inference from Utaztekan linguistic evidence. In Swanson, Earl H., Jr. (ed.), *Utaztekan prehistory*. Pocatello: Idaho State University Museum. 1-42
- Haas, Mary R. 1941. Tunica. In Boas, Franz (ed.), *Handbook of American Indian languages*. Vol. 4. New York: J. J. Augustin. 1-143.
- Haas, Mary R. 1976. The Northern California linguistic area. In Langdon, Margaret & Silver, Shirley (eds.), *Hokan studies*. The Hague: Mouton. 347-359.
- Hammarström, Harald; Forkel, Robert & Haspelmath, Martin. 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org>.
- Heine, Bernd & Song, Kyung-An 2010. On the genesis of personal pronouns: some conceptual sources. *Language and Cognition* 2. 117-147.
- Katz, Joshua 1998. *Topics in Indo-European personal pronouns*. Ph.D. dissertation. Harvard University, Cambridge, MA.
- Kroeber, Alfred L. 1913. The determination of linguistic relationship. *Anthropos* 8. 389-401.
- Manzelli, Gianguido 2015. Fra universali fonologici e ricostruzione linguistica: il caso delle nasali in proto-chibcha (America centrale e meridionale). In Busà, Maria Grazia & Gesuato, Sara (eds.), *Lingue e contesti: studi in onore di Alberto M. Mioni*. Padua: Cleup. 293-311.
- Matras, Yaron 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matteson, Esther 1972. Towards Proto Amerindian. In Matteson, Esther (ed.), *Comparative studies in Amerindian languages*. The Hague: Mouton. 21-89.
- Meira, Sérgio 2002. A first comparison of pronominal and demonstrative systems in the Cariban language family. In Crevels, Mily; van de Kerke, Simon; Meira, Sérgio & van der Voort, Hein (eds.), *Current studies on South American languages*. Leiden: Research School of Asian, African, and Amerindian Studies (CNWS). 255-275.
- Muysken, Pieter 2008. *Functional categories*. Cambridge: Cambridge University Press.
- Newman, Stanley 1979. A history of the Salish possessive and subject forms. *International Journal of American Linguistics* 45. 207-223.
- Nichols, Johanna 1996. The comparative method as heuristic. In Durie, Mark & Ross, Malcolm (eds.), *The comparative method reviewed: regularity and irregularity in language change*. New York: Oxford University Press. 39-71.
- Nichols, Johanna & Peterson, David A. 1996. The Amerind personal pronouns. *Language* 72. 336-371.
- Nichols, Johanna & Peterson, David A. 1998. Amerind personal pronouns: a reply to Campbell. *Language* 74. 605-614.
- Nichols, Johanna & Peterson, David A. 2005. Personal pronouns. In

- Haspelmath, Martin; Dryer, Matthew S.; Gill, David & Comre, Bernard (eds.), *The world atlas of language structures*. Oxford: Oxford University Press. 546-553.
- Ortman, Scott G. 2012. *Winds from the North: Tewa origins and historical archaeology*. Salt Lake City: University of Utah Press.
- Parkvall, Mikael 2008. Which parts of language are the most stable? *Sprachtypologie und Universalienforschung* 61. 234-250.
- Proulx, Paul 1985. Proto-Algic II: verbs. *International Journal of American Linguistics* 51. 59-93.
- Rankin, Robert L.; Carter, Richard T.; Jones, A. Wesley; Koontz, John E.; Rood, David S. & Hartmann, Iren (eds.) 2015. *Comparative Siouan dictionary*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://csd.cld.org>.
- Rijkhoff, Jan & Bakker, Dik 1998. Language sampling. *Language Typology* 2. 263-314.
- Ross, Malcolm 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Chang, Henry Y.; Huang, Lillian M. & Ho, Dah-an (eds.) *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*. Taipei: Institute of Linguistics, Academia Sinica. 521-564.
- Sapir, Edward 1918. Mosesteno vocabulary and treatises by Benigno Bibolotti. *International Journal of American Linguistics* 1. 183-184.
- Sasse, Hans-Jürgen 2015. Syntactic categories and subcategories. In Kiss, Tibor & Alexiadou, Artemis (eds.) *Syntax – theory and analysis*. Berlin, Munich & Boston: De Gruyter Mouton. 158-217.
- Satzinger, Helmut 2004. Statuses and cases of the Afroasiatic personal pronoun. In Takács, Gábor (ed.) *Egyptian and Semito-Hamitic (Afro-Asiatic) studies in memoriam W. Vycichl*. Leiden / Boston: Brill. 487-498.
- Sayce, Archibald H. 1874-1875. *The principles of comparative philology*. London: Trübner.
- Shaul, David L. 2014. *A prehistory of western North America: the impact of Uto-Aztecan languages*. Albuquerque: University of New Mexico Press.
- Swadesh, Morris 1954. Perspectives and problems of Amerindian comparative linguistics. *Word* 10. 306-332.
- Swadesh, Morris 1960. *Estudios sobre lengua y cultura*. Mexico City: Escuela Nacional de Antropología e Historia.
- Thomason, Sarah G. & Kaufman, Terrence 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tolmie, W. Fraser & Dawson, George M. 1884. *Comparative vocabularies of the Indian tribes of British Columbia*. Montreal: Dawson Brothers
- Trombetti, Alfredo 1905. *L'unità d'origine del linguaggio*. Bologna: Libreria Treves di Luigi Beltrami.
- Willerman, Raquel 1994. *The phonetics of pronouns: articulatory bases of markedness*. Ph.D. dissertation. University of Texas, Austin.