The millennial persistence of Indo-European and Eurasiatic pronouns and the origin of nominals

Pierre J. Bancel & Alain Matthey de l'Etang
Association for the Study of Language in Prehistory (ASLIP)

To Hal Fleming, a true mind and friend

In further defense of the Proto-Sapiens antiquity of global kinship etymologies, we illustrate the long-lasting survival of personal pronouns first in the Indo-European family, then in the Eurasiatic macrofamily of languages. We then put forward a conjecture about how the category of 1st and 2nd person pronouns might have originated.

1. Presentation

The claim that *papa, mama* and *kaka* kinship terms have been inherited since the ±100,000 year-old Proto-Sapiens origin of all human languages (Bengtson & Ruhlen 1994; Ruhlen 1994a; Bancel & Matthey de l'Etang 2002; Matthey de l'Etang & Bancel 2005) flies in the face of the widespread belief that words are replaced at such quick pace that it would erase any trace of an original vocabulary in less than 10,000 years. This belief would imply that ancestral words disappear from the lexicon of all languages at a relatively regular pace.

Such is not the case, however. Some words last much longer than others. Dolgopolsky (1964) already found that the 1st and 2nd person pronouns are the first and third most resistant word meanings, respectively. More recently, Pagel (2000, quoted in Ruhlen 2007) found comparable results and calculated that 1st person singular pronouns have an average half-life exceeding 21,000 years.

We will flesh out these statistical claims with solid linguistic facts. In the Indo-European language family, 1st person *m-* and 2nd person *t-* pronoun roots have been preserved in all descendant languages over some 7,000 to 8,000 years. Moreover, in the still more ancient Eurasiatic family, 1st person *m-* and 2nd person

t- pronouns have been preserved in most groups, and in almost all their descendant languages for at least 10,000 years, as claimed by Greenberg (2000). We conclude that Dolgopolsky and Pagel are right, and that some words may last for almost indefinite timespans. Thus, the a priori objection against the Proto-Sapiens antiquity of kinship terms, based on any average word replacement rate, simply does not hold. Finally, we propose a conjecture about the fact that, like most other ancient pronominal roots in the world's languages (Ruhlen 1994b), the *m-* and *t-* Eurasiatic pronouns are built from the same consonants as Proto-Sapiens kinship terms. In our opinion, the ancestral kinship terms have been the precursors of personal pronouns.

2. The exceptional persistence of Indo-European pronouns

Indo-European languages give us a clear example of the pronouns' exceptional resistance. Indo-European languages descend from Proto-Indo-European, a language spoken some 7,000 to 8,000 years ago. Proto-Indo-European was a language with nominal declension, marking cases with suffixes. So were all its early written descendants such as Hittite, Sanskrit, Avestan, Tocharian, Classical Armenian, Classical Latin, Classical Greek, Old Church Slavic, or Gothic, and still are many modern descendant languages such as Hindi, Modern Armenian, Rumanian, Modern Greek, Russian, German, and so on. We will examine here in full detail the 1st person pronoun only.

2.1 Indo-European 1st person pronoun

Personal pronouns took part to the nominal declension system. Among them, the 1st person pronoun had the striking peculiarity to be built on two different roots. There was an isolated form *eg^{h}om* "I" for the nominative (subject) case, alternating with a root *m-* bearing various suffixes marking the accusative, genitive, dative, and ablative cases. This use of a special word for the nominative of the 1st person singular pronoun was isolated within the Proto-Indo-European declension system. All other nouns, pronouns, and adjectives used the same root for the nominative as for all other cases, each case being marked by a specific suffix. From a functional viewpoint, the word *eg^{h}om* was doubly weird: (1) within the 1st person singular pronoun's declension, it implied to remember two entirely different word roots; (2) within the whole declension system, it implied to remember that the 1st person pronoun behaves differently from all other nominals. In an evolutionary perspective, the *eg^{h}om ~ m-* alternance was a typical candidate for an analogical simplification which would have generalized one of the two forms. We will see that this theoretical hypothesis is amply verified in the Indo-European languages' history.

Table 1. Evolution of *eg^{h}om* "I" ~ *me* "me" in languages descended from Proto-Indo-European

Italic		**Slavic**		**Armenian**	
Class. Latin	*egŏ ~ mē*	O. Ch. Slavic	*jazŭ ~ mę*	Cl. Arm.	*es ~ im* (gen.)
French	*je ~ me*	Russian	*ja ~ me(-nja)*	Mod. Arm.	*jes ~ im* (gen.)
Portuguese	*eu ~ me*	Ukrainian	*ja ~ me(-ne)*		
Spanish	*yo ~ me*	Polish	*ja ~ m(-nie)*	**Iranian**	
Catalan	*jo ~ me*	Czech	*já ~ me*	Avestan	*azəm ~ ma(-m)*
Occitan	*ieu ~ me*	Slovak	*ja ~ ma*	Farsi	*ma(-n)*
Italian	*io ~ me*	Serbo-Croat.	*ja ~ me*	Baluchi	*ma(-n)*
Rumanian	*eu ~ me*	Slovenian	*jaz ~ me*	Kurdish	*ez ~ mi(-n)*
		Bulgarian	*az ~ me*	Ossetic	*æz ~ my(-n)*
Celtic				Pashto	*zə ~ maa*
Old Irish	*mē*	**Germanic**			
Middle Welsh	*mi*	Gothic	*ik ~ mi(-k)*	**Nuristani**	
Mod. Welsh	*mi*	English	*I ~ me*	Kâmviri	*ōtθ ~ ıa*
Cornish	*my*	Dutch	*ik ~ mi(-j)*	Kâtavari	*uze ~ ie*
Breton	*me*	German	*ich ~ mi(-ch)*	Sañuvîri	*âi ~ yũ*
		Danish	*jeg ~ mi(-g)*		
Greek		Swedish	*jag ~ mi(-g)*	**Indic**	
Class. Greek	*egō(n) ~ (e-)me*	Icelandic	*ég ~ mé(-r)*	Sanskrit	*ahám ~ mā́(-m)*
Mod. Greek	*égo ~ mu*			Pali	*aham ~ ma(-m)*
		Anatolian		Hindi	*mẽ ~ m(-ujh)*
Albanian		Hittite	*ūk ~ amm(-uk)*	Bengali	*am-*
Albanian	*unë ~ më*	Luvian	*? ~ mi/mu*	Punjabi	*me(-n)*
				Gujarati	*hu(-n) ~ ma(-ri)*
Baltic		**Tocharian**		Kashmiri	*bŏh ~ me*
Old Prussian	*aš ~ ma(-nę)*	Tocharian A	*ñ-uk* f. ~ *n-ä s* . m.	Marathi	*mi ~ ma(-dža)*
Latvian	*es ~ ma(-ni)*	Tocharian B	*ñ-aś*	Romany	*me ~ ma(-n)*

The first form is that of nominative case; the second form is that of accusative, except for Armenian, where it is replaced by that of genitive, and Tocharian A, where the alternance is to be observed between the two feminine and masculine forms of nominative. Underlined languages have generalized a single root for all cases and did not preserve any trace of the ancient alternance. Elements between parentheses are part of the word in the descendant language, but have historically been added to the root by analogy. After 7,000 to 8,000 years of evolution, all thirteen Indo-European branches but Celtic have preserved the alternance between nominative and the other cases, with two forms most of the time derived from *eg^{h}om* and *m-*. All Indo-European languages, without a single exception, have preserved the non-subject root *m-*, sometimes under an evolved form but mostly unchanged. Note that in several groups the original Proto-Indo-European genitive suffix *-n* (in the PIE genitive *me-ne* "of mine") has been generalized to a great part of the declension (Russian, Latvian, Romany), or even to the whole paradigm (Farsi, Baluchi, Tocharian), or survives in the possessive (German *mein,* English *mine*).

Albeit apparently weakly functional, the two alternating forms *eg^{h}om ~ m-* have survived until the present day in an amazing number of Indo-European languages. Table 1 illustrates its preservation with representative examples of the 1st person singular pronoun in all the Indo-European linguistic groups. In Table 1, in the box facing each language name, the nominative form comes first (in Tocharian A, the first form is the feminine variant of the nominative), separated by a tilde (~) from the accusative form (or the genitive, for Armenian, or the masculine form of nominative, for Tocharian A). Thus, in Classical Latin *egŏ ~ mē, egŏ* is the nominative form ("I") and *mē* the accusative ("me"). Languages for which a single root (as Tocharian B *ñ-*) or word (as Old Irish *mē*) is given use the same root or word in all grammatical cases.

The first language listed in each group is the ancient language closest to the origin, e.g., Classical Latin for the Italic group, or Avestan for Iranian. This immediately allows to note that nearly all the ancient Indo-European languages preserved both a derivative of *eg^{h}om* in the nominative and a derivative of *m-* in the other cases: Hittite *ūk ~ amm(-uk),* Vedic Sanskrit *ahám ~ ma(-m)* (with two variants [m$^{\frac{1}{2}}$] and [m$^{\frac{1}{2}}$-m]), Avestan *azəm ~ mą(-m),* Classical Greek *egō* (variant *egōn*) *~ me,* Classical Latin *egŏ ~ mē,* Gothic *ik ~ mi(-k),* Old Church Slavic *jazŭ ~ mę,* Old Prussian *aš ~ ma(-nę),* Tocharian A *(ñ-)uk* (feminine nominative) *~ ñ-* (masculine nominative *ñ-äṣ* and all other cases), Classical Armenian *es ~ im* (genitive), etc. Only Old Irish lost any trace of *eg^{h}om,* like all other, more recent Celtic languages, and generalized a derivative of *me* in the nominative. (In the examples above, the elements between parentheses linked to the rest of the word by a hyphen are part of the form cited, but have been added to the original form in the course of history by an analogical process. For instance, the Gothic accusative *mik* "me" evolved from an original *mi* to which a final *-k* was added by analogy with nominative *ik* "I;" Venetic, an ancient Italic language, did the same with *eχo* "I" ~ *meχo* "me.")

Among the thirteen Indo-European branches, only the Celtic group did not preserve any trace of *eg^{h}om.* In this group, the root *m-* was generalized to all cases, nominative included. This analogical replacement confirms the poor functionality of using two roots for the same meaning. Table 1 offers several additional instances of this replacement in the Tocharian, Iranian, and Indic groups. In these groups, whose respective proto-languages did preserve the *eghom ~ m-* opposition, some languages – e.g., Farsi, Baluchi, Hindi, Marathi, or Romany – later lost it and replaced the form derived from *eg^{h}om* by a form derived from *m-*. The fact that these languages belong to different groups whose respective proto-languages preserved a form derived from *eg^{h}om* implies that this analogical simplification occurred several times independently. To these languages may be added the case of French, a Romance language in which the descendant form *je* of Latin *egŏ* has

become close to a verbal prefix marking a 1st person subject, while the subject form of the independent pronoun is now *moi,* derived from a stressed variant of Latin *mē*.

It is thus a heavy trend from the part of the speakers of any inflected language to use a single root for all the uses of a word. However, the career of the two alternating forms of Proto-Indo-European 1st person pronoun did not stop with the ancient languages descended from Proto-Indo-European. Three millennia later, they continue to coexist in a very high number of modern languages. In spite of the restriction mentioned above, Modern French preserves both a derivative *je* from Latin nominative *egŏ* and derivatives *me, moi* from Latin accusative *mē*. Indeed, Modern French grammar abandoned the whole declension system inherited from Latin except in pronouns. The use of different forms according to whether the personal pronoun is subject (*Je vois l'homme* "I see the man") or object (*L'homme me voit* "The man sees me") practically constitutes in French the only trace (with the relative pronoun) of the declension system inherited by Latin from Proto-Indo-European.

The loss of declension in Modern French makes the preservation of the opposition between *je* and *me* ~ *moi* still more amazing. On the one hand, French preserves an atypical and – in terms of mnemonic efforts – costly opposition between two roots bearing the same meaning (which several other Indo-European languages have independently reduced to a single root); on the other hand, by doing so, French preserves a trace of a grammatical system having otherwise entirely disappeared. As paradoxically as French, all other Romance languages do preserve the two forms, together with their grammatical distribution between subject and non-subject, although almost all of these languages – to the single exception of Rumanian – otherwise lost, like French, the ancestral declension system.

Beyond Romance languages, the two terms inherited from Proto-Indo-European have been preserved, in their respective original grammatical functions, in all Germanic languages (even though Modern English lost, like French, any trace of declension other than in pronouns) as well as in all Baltic and Slavic languages, Modern Greek and Modern Armenian. Modern Albanian also maintains the opposition between subject and non-subject forms; however, while it preserved the original *m-* for the latter, it is unsure whether the subject form *unë* contains a derivative of *eg^{h}om*.

A number of modern Indic languages descended from Sanskrit did, like Celtic, generalize the non-subject *m-* to all cases: Hindi and Urdu *mẽ,* Punjabi *me(-n),* Marathi *mi,* etc. However, Gujarati as well as a Dardic language like Kashmiri still preserve the root alternation between subject and non-subject forms (but they replaced the derivative of the Sanskrit subject form *ahám* by new forms).

Nuristani languages (Strand 1997–2008) preserved the alternance between subject and non-subject forms. Kâtavari and Kâmviri subject forms *uze* and *ōtθ* "I" clearly descend from Proto-Indo-European *eg^{h}om.* Nuristani non-subject forms have undergone the strongest phonetic erosion, to the point of hardly preserving a trace of the original *m-* as a nasalization of the neighboring vowel, as in Kâmviri *īa* "me" (but Kâtavari *ia* "me" even lost this nasal feature); however, the original *m-* survives in the 1st person singular possessive pronoun (Kâtavari *iema,* Sañuvîri *imâ* "my") as well as in the 1st person plural personal pronoun (Kâmviri *imo,* Kâtavari *imu,* Sañuvîri *ima* "we"). In the Iranian group, Farsi and Baluchi *ma(-n)* only preserve the non-subject form, generalized to all cases. To the contrary, Kurdish *ez ~ mi(-n)* preserves the original opposition, and so do Ossetic *æz ~ my(-n)* and Pashto *zə ~ maa.*

Moreover, *m-* has also been preserved in numerous 1st person plural pronouns (e.g., Russian *my,* Lithuanian *mes,* Armenian *mek‘* "we"), 1st person singular possessives (e.g., French *mon,* English *my, mine,* German *mein,* Russian *moj,* Hindi *mera,* "my"), and 1st person verb suffixes (e.g., Latin *sum* "I am," *sumus* "we are," *amabam* "I loved," *fuimus* "we were," French *nous allons* "we go," *nous fûmes* "we were," Classical Greek *eimi* "I am," *pheromai* "I bear for myself," *pheromen* "we bear," Irish *bim* "I am," Albanian *unë jam* "I am," *ne jemu* "we are," Latvian *esmu* "I am," Russian *ja dam* "I will give," *my berjom* "we bear," Armenian *berem* "I bear," Tocharian A *nas-am* "I am," *nas-amäs* "we are," Avestan *barāmi* "I bear," Sanskrit *bhar-āmi* "I bear," etc.). Finally, let us observe that the *-n* suffix of Proto-Indo-European genitive *mene* "of mine" was extended to the accusative and other non-subject cases in several Indo-European groups (e.g., Russian *menja,* Latvian *mani,* Kurdish *min* "me") and even generalized to the nominative (Tocharian B *ñaś*, Farsi *man,* Punjabi *men* "I").

In sum, the reality appearing from Table 1 is that all modern Indo-European languages – from Icelandic to Russian and Bengali through Welsh, Armenian, Greek, Kâmviri and Pashto) – ALL, WITHOUT A SINGLE EXCEPTION, preserve a derivative from the original *m-*. In a small number of cases, the *m-* survives under a phonetically evolved form where it is impossible to recognize it at first glance, for instance in Tocharian B *ñ-* or still more in the Armenian accusative *is*, as well as in some Nuristani languages. All Indo-Europeanists, however, agree that these forms descend from the original Proto-Indo-European *m-*. In some languages, like Tocharian B, Farsi, Hindi and the whole Celtic group, the original *m-* was not only preserved but analogically extended to the nominative, where it replaced the derivative of *eg^{h}om.* But in a great majority of languages, Proto-Indo-European *m-* did not change at all – neither in its phonetic form, nor in its grammatical functions, nor in its "1st person" meaning.

As to the nominative form *eg^{h}om,* it survived in at least 11 groups out of 13 (to the exceptions of Celtic and perhaps Albanian). In 8 out of these 11 groups (namely, Italic, Greek, Germanic, Baltic, Slavic, Armenian, Nuristani and Anatolian), it even survived

in all their member languages; in the 3 other groups (Tocharian, Iranian, and Indic), only part of the descendant languages preserved a derivative of *eg^{h}om.* In all of the numerous languages where a derivative of *eg^{h}om* was preserved, it continues to coexist along with a derivative of *m-* (of course, since *m-* was preserved in all languages). Furthermore, the original distribution of *eg^{h}om* and *m-* between subject and non-subject cases has remained the same (or, in a few cases, almost the same). In 7,000 to 8,000 years, the replacement rate of 1st person pronoun *m-* amounts to exactly 0%, and its preservation to 100%. As to the form *eg^{h}om,* if we consider groups, the loss is of only 2/13, or 15.4%; if we consider individual languages, the fair number of Indic languages which lost it would perhaps raise this percentage to about 20% to 30%.

According to the average 14% replacement rate per millennium calculated by Swadesh (1952, 1954), pronouns should have been preserved in only 34.8% (after 7,000 years) to 29.9% (after 8,000 years) of languages descended from Proto-Indo-European. The 100% preservation of *m-* and the 70% to 84.6% preservation of *eg^{h}om* make clear that an average loss rate is meaningless with regard to the 1st person pronoun.

2.2 Indo-European 2nd person pronoun

This exceptional resistance of 1st person pronoun is also observed in the Proto-Indo-European 2nd person pronoun *t-*. This pronoun did not display any root alternation between the nominative and other cases. Table 2 displays a few representative modern forms derived from its subject form *tu* in each Indo-European group.

Table 2. Representative forms of 2nd person singular pronoun in the nominative case in languages descended from Proto-Indo-European

Groups	Ancient languages	Modern languages
Italic	Cl. Latin *tū*	Spanish *tu,* French *tu,* Italian *tu,* Rumanian *tu*
Celtic	Old Irish *tū*	Welsh *ti,* Breton *te,* Modern Irish *ti*
Greek	Cl. Greek *sú* < *tú*	Modern Greek *(és-)si* < *sú*
Baltic	Old Prussian *tū*	Lithuanian *tú,* Latvian *tu*
Slavic	Old Church Slavic *ty*	Russian *ty,* Slovak *ty,* Bulgarian *ti,* Slovenian *ti*
Germanic	Gothic *θu*	Dutch *du,* English *thou,* German *du,* Icelandic *ðú*
Albanian	—	Albanian *ti*
Armenian	Class. Armenian *dow*	Modern Armenian *tun*
Anatolian	Hittite *zik* < *ti-uk* (accusative *tuk*)	
Tocharian	Tocharian A *tu* ~ Tocharian B *tuwe*	
Iranian	Avestan *tū*	Farsi *to,* Baluchi *tao,* Ossetic *dy,* Kurdish *tu,* Pashto *tə*
Nuristani	—	Kâtavari *t'ü*, Kâmviri *t'ü,* Sañuviri *tū*
Indic	Sanskrit *tw(-əm)*	Hindi *tū,* Kashmiri *ts(-ah),* Gujarati *tu,* Oriya *tu(-me)*

All Indo-European languages preserved a 2nd person singular pronoun inherited from Proto-Indo-European *tu.*

Just like for 1st person *m-*, not a single Indo-European language did lose the Proto-Indo-European 2nd person *t-*. And most of them preserved, like for 1st person *m-*, the original consonant *t-* either unchanged or only slightly modified. Second person *t-* is also just as general as 1st person *m-* in possessives and verbal markers. Indeed, mentioning and analyzing all 1st person *m-* forms and all 2nd person *t-* forms in Indo-European languages would take a whole book – for each.

2.3 Indo-European numerals

A few other Proto-Indo-European words also survived in nearly all of the modern descendant languages, such as the basic numerals from "two" to "ten." Examples of derivatives of *dṷou* "two," taken from all thirteen Indo-European branches, are given in Table 3. As for 1st person *m-* and 2nd person *t-*, *dṷou* "two" survived in all and every descendant language of Proto-Indo-European – see the nice online compilation of the first ten numerals in over 4,500 languages by Mark Rosenfelder (no date).

Table 3. Representative forms of Proto-Indo-European *dṷou* "two" in descendant languages

Groups	Ancient languages	Modern languages
Italic	Cl. Latin *dúo*	Spanish *dos,* French *deux,* Italian *due,* Rumanian *doi*
Celtic	Old Irish *deu*	Breton *daou,* Modern Irish *dhaü*
Greek	Cl. Greek *dúo* ~ *dúō*	Modern Greek *ðío*
Baltic	Old Prussian *dwai*	Lithuanian *du,* Latvian *divi*
Slavic	Old Church Slavic *dŭva*	Russian *dva,* Polish *dwa,* Slovenian *dva*
Germanic	Gothic *twai*	Old Norse *tveir,* English *two,* German *zwo* ~ *zwei,* Danish *to*
Albanian	Old Albanian *do*	Tosk *dy*
Armenian	Class. Armenian *erk'u*	Modern Armenian *yergu*
Anatolian	Hittite *da*, Lycian *tuwa*, Luwian *duwa*	
Tocharian	Tocharian A *wu* ~ Tocharian B *wi*	
Iranian	Avestan *dva*	Farsi *do,* Ossetic *dyuö,* Kurdish *du,* Pashto *dva*
Nuristani	—	Kâtavari *d'ü,* Kâmviri *d'ü,* Wasiweri *lüü*
Indic	Sanskrit *dwau*	Hindi *do,* Kashmiri *zăh,* Marathi *don,* Oriya *dui*

All Indo-European languages preserved a form inherited from Proto-Indo-European *dṷou* "two."

In this series, the Armenian forms seem unrelated to the other words, but it is a well-known achievement of linguistic reconstruction to have shown that Classical Armenian *erk'u* "two" and Modern Armenian *yergu* result from a regular phonetic evolution of Proto-Indo-European *dṷou* "two."

2.4 Why persistent words, and what do they mean for language history?

A comparably high preservation rate of personal pronouns is encountered in most language families. This extraordinary persistence may be explained in a simple

word: FREQUENCY. Personal pronouns – and more generally person markers – are among the most frequent meaningful words in all languages: in the everyday discourse, most sentences include 1st and/or 2nd person markers.

This resistance to linguistic erosion gives pronouns a very high classificatory value, well known to long range comparatists. A recent attempt to invalidate this classificatory value of pronouns, explicitly directed against long-range comparison, claimed that there are languages where personal pronouns are "freely borrowed" from other languages (Thomason & Everett 2005). Their world tour in search of such cases resulted in less than two dozen languages – out of some 7,000 – in which they found examples apparently supporting this claim. Among them, about a half concern 3rd person pronouns – which are different in nature, as was shown long ago by Benveniste (1946). Finally, among their examples concerning 1st and 2nd person pronouns, several are more than dubious, e.g., when they ascribe the striking convergence between 1st person pronouns in Turkic, Mongolic and Tungusic languages (see sections 3.2 to 3.4 below) to borrowing rather than to common inheritance from Proto-Altaic – while they admit overtly that they "have not carried out any kind of systematic analysis of the Korean, Japanese, Tungusic, Mongolian or Turkic data."

On the basis of such scanty and uncertain evidence, the authors of this study go so far as concluding that "we can't know […] whether a given past culture was more like Indo-European languages, where most pronouns are in fact inherited, or more like those of languages whose speakers borrow pronouns freely." In fact, the weakness of their study *a contrario* demonstrates that pronoun borrowing is really exceptional, unintentionally confirming the classificatory value of 1st and 2nd person pronouns.

Less exceptional is pronoun replacement WITHIN a given language, due to a specific cultural attitude, leading for instance to replace the 2nd person singular pronoun by a plural or a 3rd person as a mark of respect (English lost its 2nd person singular by borrowing from Old French, if not the pronoun itself, the use of the 2nd plural when speaking to a single addressee), or tabooing the use of 1st person singular pronoun as an impolite expression of selfishness. It is worth noting, by the way, that such a taboo may have been at the origin of the Indo-European suppletion between the nominative and other cases in the 1st person pronoun: the taboo would have borne on the subject case, i.e., in sentences where the speaker directly refers to his own actions or feelings, but not on non-subject cases, where syntax makes the role of the speaker more incidental in the action or state described.

In contrast, most other Proto-Indo-European words known to us – even other items from the Swadesh's 200 basic words list – have been preserved in much less numerous modern languages, often in only two or three branches – and, within these branches, in only some languages. Contrary to numerals and personal pronouns, such words underwent a much higher replacement rate than 14% per millennium.

Consequently, it cannot be assumed that an average 14% replacement rate per millennium does lead, after 1,000 years, to the loss of any one word out of seven or so from the ancestral vocabulary.

2.5 Personal pronouns in Eurasiatic

It is thus an established fact that Indo-European personal pronouns *m-* and *t-* have been preserved for some 8,000 years in all the numerous descendant languages of the family. If these forms have been preserved in several hundreds of languages over such a long timespan, it might well happen that they had not been new in Proto-Indo-European, but had been inherited from a remoter ancestor language. This ancestral language, of course, could have other descendants than Indo-European. And, given the pronouns' extraordinary persistence, it seems worth trying and finding their trace in the possible sister languages of Indo-European as well.

The Eurasiatic hypothesis (Greenberg 2000, 2002) links Indo-European to six other language families, namely Uralic-Yukaghir, Altaic, Nippo-Koreo-Ainu, Gilyak, Chukchi-Kamchadal, and Eskimo-Aleut. (To these seven families, Greenberg tentatively adds Etruscan.) Sixty-nine common grammatical forms (Greenberg 2000) and some 430 common lexical roots (Greenberg 2002) support the Eurasiatic unity. Among them, 1st person *m-* and 2nd person *t-* are among the most widely represented grammatical roots within the Eurasiatic family. Most of the data mentioned below about these two pronouns in the various Eurasiatic groups are given by Greenberg (2000) in his Eurasiatic grammatical etymologies (n° 1, 1st person *m*; n° 4, 2nd person *t*; n° 7, pronoun base *ge*; n° 17, plural *r(i)*; and n° 25, genitive *n*). Additional data have been taken from Starostin *et al.* (2003; no date), Bomhard (2007), and a range of descriptive works. Most analyses presented here are due to Greenberg – our contribution is essentially to synthesize his observations on Eurasiatic pronouns, which appear in his book dispersed over several sections.

All these families have a 1st person pronoun *m-*, and in four of them (Uralic-Yukaghir, Altaic, Chukchi-Kamchadal, and Eskimo-Aleut) it is not only present but also a canonical marker for 1st person independent pronouns, possessives pronouns, and verb markers, just like in Indo-European. Moreover, in six families (the same four plus Nippo-Koreo-Ainu and Etruscan), we find both a root alternation between the nominative and other cases, and an extension *-n* which was more or less generalized to non-subject cases (and sometimes also to the nominative), but which initially marked the genitive case, as appears from the recurrence, throughout these families, of a *mVn-* 1st person possessive (ultimately derived from the genitive of the independent pronoun, like English *(of) mine* or German *mein* derive from the Proto-Indo-European genitive *mene* "of me").

As to the Indo-European nominative suppletive form *eg^{h}om,* it is less widespread, since several groups alternating two forms of the 1st person pronoun have another root in the nominative case. Clear cognates of *eg^{h}om* are however found throughout Chukchi-Kamchadal (with the most salient attestations in Chukchi), together with somewhat less close parallels in several Uralic languages.

The only language in which the Eurasiatic 1st person marker *m-* seems completely lacking is Ainu. It is not a surprise, since Ainu is certainly the most diverging language within Eurasiatic. Even within the Nippo-Koreo-Ainu family, Ainu is so divergent from both Japanese and Korean that most comparatists – and even many long-rangers – do not agree that it is related to either of these two languages, nor to anything resembling the Eurasiatic macrofamily.

The 2nd person marker *t-* is almost as widespread in Eurasiatic as is the 1st person *m-*, being a canonical marker in Uralic-Yukaghir, the Mongolic branch of Altaic (plus clear traces in Tungusic), Chukchi-Kamchadal, and Eskimo-Aleut. It however seems to be lacking in the Turkic branch of Altaic, the whole Nippo-Koreo-Ainu family, and Gilyak. Once again, the specific Indo-European form *tu* in the nominative case finds a clear parallel in Chukchi-Kamchadal, with the pervasive *turi, tuze, tuza* "you" plural forms in both the Chukchi and Kamchadal branches, and most strikingly, in the latter branch with three different attestations *tu ~ tue ~ tua* "thou" from various extinct Itelmen dialects independently collected in the course of the XIXth century by American and Russian explorers.

Table 4. 1st person *m-* and 2nd person *t-* in Uralic-Yukaghir personal pronouns

Language	1st SG.	1st PL.	2nd SG.	2nd PL.
Proto-Uralic	*mɤ*	*me*	*tɤ*	*te*
Estonian	*ma ~ mina min-*	*me*	*sa* (< *ta*)	*te*
Finnish	*mä ~ minä min-*	*me*	*sinä* (< *tinä*)	*te*
Sami	*mun-*	*mí*	*dun*	*dí*
Komi	*me menõ*	*mi*	*te tenõ*	*ti*
Udmurt	*mon-*	*mi*	*ton*	*ti*
Hungarian	*én* (< *émn*) *engem(-et)*	*mi*	*te teged(-et)*	*ti*
Enets	*mod'i*	*mod'iʔ*	*tod'i*	*tod'iʔ*
Nganasan	*mannaŋ*	*meeŋ*	*tannaŋ*	*teeŋ*
Yukaghir	*met*	*mit*	*tet*	*tit*

2.6 Personal pronouns in Uralic-Yukaghir

In all Uralic languages as well as in Yukaghir, the 1st person pronoun is built on the root *m-* and the 2nd person pronoun is built on the root *t-* (Table 4). For the 1st person,

the non-subject suffix *-n* is present in most languages from all the branches of Uralic, where it displays a strong tendency to generalization to the nominative (e.g., Finnish, Ostyak), and sometimes to the 2nd person as well (e.g., Sami, Udmurt, or Nganasan). The Proto-Uralic 1st person singular subject marker on verbs is *-m* (plural *-mek*), and the 2nd person is *-t* (plural *-tek*); the 1st and 2nd person singular possessors are marked by the same forms suffixed to nouns.

Moreover, Hungarian offers an accusative form strongly reminiscent of Proto-Indo-European *eg^{h}om* "I," analyzed by Greenberg (2000) as a compound *e-g^{h}o-m* "it is me" in which the final *-m* is the 1st person marker. In Hungarian, the form would have been turned into a redundant *en-ge-m(-et)* "me that's me" (the final *-et* marks the accusative case), an analysis confirmed by the Hungarian 2nd person pronoun accusative *te-ge-d(-et),* where both initial *te-* and final *-d* clearly derive from 2nd person *t-*. Other Uralic forms such as Vogul *am-kke-m* "I alone," the Selkup accusative *šim* (< *kim*) "me" and maybe the Kamassian suppletive form of the verb "to be" in the present tense (1st person *i-gä-m* "I am") may belong to the series, which finds a clear parallel in Chukchi-Kamchadal (see Table 10 below).

2.7 Personal pronouns in the Turkic branch of Altaic

The Turkic branch of Altaic is divided into two main groups, Chuvash and non-Chuvash. Chuvash offers an alternation between nominative *e-pə* and the non-subject cases built on the root *m-* plus the non-subject suffix *-n*, giving in Chuvash *man-* (Table 5). Chuvash, the most divergent Turkic language, thus parallels exactly the general state of things for the 1st person pronoun within the two other Altaic branches, Mongolic (see Table 6) and Tungusic (see Table 7).

Table 5. 1st person *m-* in Turkic personal pronouns (no 2nd person *t-*)

Language	1st SG.	1st PL.	Language	1st SG.	1st PL.
Chuvash	*e-pə man-*	*e-pir pir*	**Azeri**	*män-*	*biz-*
Old Turkish	*ben- ~ men-*	*biz-*	**Mod. Turkish**	*ben-*	*biz-*
Uzbek	*men-*	*biz-*	**Gagauz**	*ben-*	*bis-*
Yakut	*min-*	*bihigi*	**Kazakh**	*min-*	*biz-*
Shor	*men-*	*pis-*	**Kirghiz**	*men-*	*miz-*
Oyrot	*men-*	*mis-*	**Tatar**	*min-*	*bez-*

In the other Turkic group, the Old Turkish free alternation *ben- ~ men-* in the 1st person singular illustrates the first step in the reinterpretation of the formerly grammatically distributed variants, together with the generalization of the non-subject suffix

-n to the whole singular paradigm. In present-day non-Chuvash Turkic languages, the two variants *b-* and *m-* are variously distributed between singular and plural, according to languages: Modern Turkish generalized *b-* to both singular and plural (where suffix *-n* is replaced by the plural suffix *-iz* > *-ri*, an alternation general in Turkic); conversely, Oyrot, Kirghiz and Balkar generalized *m-* to both singular and plural; many languages, such as Tatar, Azeri, Khalaj, Yakut, or Nogai, generalized *m-* in the singular pronoun and *b-* in the plural. As to the former non-subject suffix *-n,* it was already generalized to all singular cases in Old Turkish and is retained until today by all non-Chuvash languages as the mark of the singular in the 1st person pronoun. 2nd person marker *t-* seems to be lacking in Turkic, as the canonical *sen* "thou" ~ *siz* "you" are not thought to have evolved from *t-* forms. The non-subject suffix *-n,* which was generalized in the 1st person singular, appears in the 2nd person singular pronoun as well, like in many Uralic languages.

2.8 Personal pronouns in the Mongolic branch of Altaic

In Mongolic, the alternation between the suppletive 1st person *b-* in the nominative and *mVn-* in other cases was extended to the plural (Table 6). Once again, we retrieve the same configuration as for the Indo-European 1st person, with a root alternation opposing the nominative to other cases, with a root *m-* and a suffix *-n* in the genitive that was generalized in Mongolic to both the singular and plural of the 1st and 2nd persons. The only difference is that, instead of *eg^{h}om* for the suppletive nominative form, we have a nominative root in *b-*, like in Chuvash. It is certainly interesting to observe that, phonologically, this *b-* is the non-nasal counterpart of the non-subject root *m-*.

Table 6. 1st person *m-* and 2nd person *t-* in Mongolic personal pronouns

Language	1st SG.		1st PL. EXCL.		2nd SG.	2nd PL.		1nd PL. INCL.	
Mid. Mongolian	*bi*	*mino*	*ba*	*mano*	*či čino*	*ta*	*tan-*	*bida*	*bidan-*
Class. Mong.	*bi*	*minu*	*ba*	*man-*	*či činü*	*ta*	*tan-*	*bida*	*bidan-*
Khalkha	*bi*	*minii*	*biddə*	*man-*	*či činii*	*tā*	*tan-*	*biddə*	*biddn-*
Ordos	*bi*	*mini*	*bida*	*man-*	*či čini*	*ta*	*tan-*	*bidə*	*bidn-*
Kalmuk	*bi*	*minē*	*bidə*	*man-*	*či činə*	*ta*	*tan-*	*bidʌ́*	*bidn-*
Buriat	*bi*	*menī*	*bɛde*	*man-*	*ši šenī*	*tā*	*tan-*	*bɛde*	*man-*
Dagur	*bī*	*minī*	*bā*	*mān-*	*ši šinī*	*tā*	*tan-*	*beda*	*bedańī*
Monguor	*bu*	*muni*	*buda*	*ndāni*	*ći ćini*	*ta*	*tan-*	*buda*	*budasgi-*
Paoan	*be*	*mene*	*mange*	*mane*	*čə čənə*	*ta*	*tan-*	*bede*	*bedan-*
Dongxiang	*bi*	*mini*	*bidžien*	*ma-*	*čə čəni*	*ta*	*tan-*	*matan*	*matan-*
Moghol	*bi*	*mini*	*(-u̯)*	*mōn-*	*či činɛi*	*to*	*ton-*	*bidä*	*mōn-*

The form on the left is the nominative, that on the right is the genitive.

As to the 2nd person, the alternation opposing *č* in the singular to *t-* in the plural is obviously a consequence of the following vowel, the singular *-i* having transformed a former *t-* into *č* – a well-known and widespread effect of closed front vowels like *i*. Finally, the 2nd person *t-* is certainly present in the Mongolic 1st person plural inclusive – i.e., encompassing the speaker and the hearer(s) – *bi-da* "we [= I + thou/you]", a clear compound of the 1st person singular *bi-* "I" and a 2nd person plural pronoun *-ta*.

2.9 Personal pronouns in the Tungusic branch of Altaic

In Tungusic (Table 7), the 1st person pronoun follows the same pattern already observed in Chuvash and the whole Mongolic branch, which also left clear traces in the non-Chuvash group of Turkic. The nominative root consonant is *b-*, and the non-subject cases are built on *mV-n-*. Like in Mongolic, this pattern was generalized to the 1st person plural pronoun.

Table 7. 1st person *m-* and 2nd person *t-* in Tungusic personal pronouns

Language	1st SG.		1st PL.		2nd SG.	2nd PL.	1st PL. INCL.
Even	*bi*	*min-*	*bu*	*mun-*	—	—	*mut*
Evenki	*bi*	*min-*	*bu*	*mun-*	—	—	*mit ~ bit*
Negidal	*bi*	*min-*	*bu*	*mun-*	—	—	*bitta*
Solon	*bi*	*min-*	*bū*	*mun-*	—	—	*miti*
Nanai	*bi ~ mi*	*min-*	*bū ~ mu*	*mun-*	—	—	*bue*
Udehe	*bi*	*min-*	*bu*	*mun-*	—	—	*minti*
Jürchen		*min-*		—	—	—	—
Manchu	*bi*	*min-*	*be*	*men-*	—	—	*muse*

The root consonant *t-* does not appear in the 2nd person pronouns, but, as in Mongolic, it is clearly seen in the compounded 1st person inclusive pronoun.

2.10 Personal pronouns in Nippo-Koreo-Ainu

The Nippo-Koreo-Ainu taxon, posited by Greenberg as one of the seven primary Eurasiatic groups, is certainly the most disputed of all seven (Table 8). The relation between Korean and Japanese is popular among long-rangers, who often posit a relation of the two with Altaic – an idea dismissed by Greenberg as a geographical artifact. Ainu is still considered as an isolate by traditional linguists, and several long-rangers link it to Austric rather than to Eurasiatic. Pronouns illustrate these divergences, showing that these three languages, and particularly Ainu, are highly divergent within Eurasiatic.

Table 8. 1st person *m-* in Nippo-Koreo-Ainu personal pronouns

Language	1st SG.		1st PL.
Modern Japanese	*wa*		—
Archaic Japanese	*wa*	*wanu*	—
Ryukyuan	*baa*	*banu*	—
Korean	—		? *uli* (if *uli* < *wuli* < *buli*)
Ainu	—		—

Only Japanese, and particularly the quite specific Ryukyuan dialects, display a 1st person pronoun reminding a Eurasiatic pattern, with its root *b-* in the nominative to which adds an extension *-n* in the accusative. It is closely parallel to the Altaic pattern, except that the root alternation *b-* ~ *m-* is not present, a fact which could easily be explained by an analogical simplification of the kind observed in several Indo-European branches and in non-Chuvash Turkic. The same simplification also occurred in Uralic – except that, in this latter family, it is the *m-* which was generalized to the whole paradigm. A single attestation *wanu* "me," from an undetermined Archaic Japanese dialect, parallels this pattern with *wa* (subject) ~ *wanu* (non-subject), though its root consonant must have undergone a classical phonetic evolution *b* (> *v*) > *w.*

With *uli* "we," possibly descended from an earlier and unattested *wuli* > *buli,* Korean also presents a possible trace of a Eurasiatic 1st person plural pronoun. Ainu seems to lack any trace of a pronoun relatable to the 1st person pronoun *m-* or *b-* series, nor does any language in the Nippo-Korean-Ainu group display anything like a 2nd person pronoun *t-*.

2.11 Personal pronouns in Gilyak

Gilyak, traditionally considered as a language isolate, displays a clear trace of 1st person *m-* only in the dual and plural forms (Table 9). In the dual *megi* "the two of us," *-gi* is the canonical Gilyak dual suffix, which reflects a Eurasiatic dual marker *ki* (Greenberg 2000, etymology 14), also attested in Uralic (and perhaps Yukaghir), Turkic, Ainu, Chukchi-Kamchadal and Eskimo, as well as possibly in Indo-European.

Table 9. 1st person *m-* and 2nd person *t-* in Gilyak personal pronouns

Language	1st SG.	1st dual	1st PL.	2nd SG.	2nd PL.
Gilyak	—	*megi*	*mer*	*či* ~ (*ti*)	—

In the plural *me-r* appears the Eurasiatic plural marker *-ri,* also present in Indo-European (e.g., Latin *ama-mu-r* "we are loved") and in the plural of Turkic pronouns, as well as in Chukchi-Kamchadal plural pronouns, e.g., Chukchi *mu-ri* "we," *tu-ri*

"you;" possible parallels appear in Tungusic, Korean, and Ryukyuan (Greenberg 2000, etymology 17). The 2nd person *t-* appears in a single attestation from a XIXth century description giving *ti* "thou" instead of the usual Gilyak pronoun *či.*

2.12 Personal pronouns in Chukchi-Kamchadal

In the Chukchi-Kamchadal family, 1st person *m-* and 2nd person *t-* are present in both the singular and plural pronouns, with very different respective patterns in the singular and the plural (Table 10). In the singular, they are preceded by a prefix *gə-,* giving e.g., Chukchi *ɣə-m* "I," *ɣə-t* "thou," or Koryak Palana *ɣə-mme* "I," *ɣə-tte* "thou." In Western Itelmen *kə-mma* "I," *kə-zza* "thou," the 2nd person marker *t-* changed to *z-,* through phonetic evolution or under influence of the plurals *mu-za* "we," *tu-za* "you."

Table 10. 1st person *m-* and 2nd person *t-* in Chukchi-Kamchadal personal pronouns

Language	1st SG.		1st PL.		2nd SG.		2nd PL.	
Chukchi	*(e)ɣəm*	*ɣəmnin*	*muri*	*muryin*	*(e)ɣət*	*ɣənin*	*turi*	*turyin*
Palana	*ɣəmme*	*ɣəmnan*	*muri*	*moryinan*	*ɣətte*	*ynanna*	*turi*	*toryinan*
W. Itelmen	*kəmma*	*kəmma'n*	*muza*	*muza'n*	*kəzza*	*kəzza'n*	*tuza*	*tuza'n*
Sth. Itelmen	*kim*	*ma*	*muš*	*burin*				
Nth. Itelmen			*buže*					
N.E. Itelmen					*tu*			
E. Itelmen			*buze*		*tue*		*suze*	
Itelmen X					*tua*			

This pattern parallels Indo-European *eg^{h}om* "I," and Chukchi offers an exact correspondent of Indo-European *eg^{h}om* in the suffixed predicative form of singular pronouns: X-*eyəm* "I am X," X-*eyət* "thou art X" (litterally, "X-I" and "X-thou"), suffixed to both nouns (e.g., *ənpənačy-iyəm* "I am an old man" [litt. "old man-I"]) and stative verbal forms (e.g., *yə-čejv-iyəm* "I am gone" [litt. "gone-I"]; the variant *-iyəm* results from vowel harmony). Greenberg analyzes these forms (and those already seen in Uralic) as compounds of a demonstrative *e-* plus a "pronoun base" *-ge-* used as copula plus 1st person *- m.* In our opinion, the pronoun base *-ge-* is linked to the Eurasiatic comitative *ko(n) ~ ko(m)* (Greenberg's etymology 34, strikingly represented essentially in Indo-European and Chukchi-Kamchadal).

The plural pronouns are built without prefix, but receive a suffixed plural marker, witness the Chukchi opposition *mu-ri* "we," *tu-ri* "you," paralleled by Western Itelmen *mu-za* "we," *tu-za* "you." Greenberg quotes XVIIIth and XIXth centuries sources of particular interest, which describe three different extinct Itelmen dialects with an *m- ~ b-* alternation (strongly reminding the Altaic alternation) distributed over the subject and non-subject forms of the 1st person plural, e.g., Southern Itelmen *mu-š*

"we (subj.)," *bu-ri-n* "our" (literally "of us"), where appear both the *-ri* plural suffix already found in Latin, Turkic, Korean and Gilyak, and the now familiar non-subject suffix *-n*. This non-subject suffix *-n* is general in all the singular and plural non-subject forms of Chukchi-Kamchadal pronouns.

As observed by Greenberg, Chukchi-Kamchadal is with Indo-European the only Eurasiatic group attesting a vowel *u* for the 2nd person pronoun, which was apparently generalized to the 1st person plural. In all surviving Chukchi-Kamchadal languages, vowel *u* was lost in the 2nd person singular, probably as a result of the generalization of the *yə-t* "thou" pattern. However, ancient independent sources once again give 2nd person singular forms *tu-* "thou" for three extinct Itelmen dialects, thus establishing another close parallel between Chukchi-Kamchadal and Indo-European. The Chukchi 2nd person plural subject marker is *-təkˌ* just like in Uralic. Finally, non-subject forms of both singular and plural 1st and 2nd person pronouns regularly display a suffix *-nˌ* which also appears in all Chukchi-Kamchadal possessives, strongly contributing to establish its original genitive function.

2.13 Personal pronouns in Eskimo-Aleut

In the Eskimo branch, the 1st person singular and plural pronouns display an inter-dialectal variation between a form in *m-* in Sirenik, an Eskimo language spoken in Siberia and probably constituting a primary branch of Eskimo, together with Yupik and Inupik (unless it is a primary branch of Yupik), and a *w-* ~ *v-* form in Yupik and Inupik (Table 11). This alternation reminds the *m-* ~ *b-* alternation found throughout Altaic and in some Itelmen dialects.

Table 11. 1st person *m-* and 2nd person *t-* in Eskimo-Aleut personal pronouns

Language	1st SG.	1st PL.	2nd SG.	2nd PL.
Aleut	—	*-man/-mas*	—	—
Sirenik	*məŋa*	*məkəta*	*əɬpi*	*ɬpəsi*
Chugach	*xwi*	*xwaŋkuta*	*äɬpäk*	*äɬpici*
Central Alaskan Yupik	*wīī wīīŋa*	*waŋkuta*	*əɬpət*	*əɬpəci*
North Alaskan Inupik	*uvaŋa*	*uvaɣt*	*ilvic*	*iɬifsi*
East Canadian Inupik	*uvaŋa*	*uvaɣt*	*ivvit*	*ilitsi*
Greenlandic	*uaŋa*	*uaɣt*	*iɬɬit*	*iɬɬissi*

In different Yupik dialects an alternation between a naked subject form and a non-subject form with a nasal suffix is observed, e.g., Central Alaskan Yupik *wī* ~ *wīŋa*. Although this nasal is *-ŋ* instead of the expected *-nˌ* the alternation is quite close to the usual Eurasiatic pattern. In Aleut, the 1st person *m-* is present only in the suffixed plural forms *-man* (Eastern and Western Aleut) or *-mas* (Central Aleut).

The 2nd person *t-* is not as obviously represented in Eskimo personal pronouns, but specialists in Eskimo count the final consonant *-t ~ -n* of *ǝɬpǝt~ ilvin* as the root of this pronoun. The interdialectal oral ~ nasal alternation in word-final stops is a well-known feature of Eskimo languages, extending well beyond the scope of personal pronouns (and occurring among labial and velar stops as well). More direct is the 2nd person singular possessive suffix *-t* (in western Eskimo languages *-n*); the 2nd person singular subject marker of intransitive verbs is *-tit,* dual *-tik* (with the dual suffix *-ki* already found in Gilyak *me-gi* "the two of us," and observed with a plural value in the 2nd person plural verbal markers of Uralic *-tek* and Chukchi *-tək*). In Aleut, the 2nd person singular possessive *-n* regularly corresponds to Eskimo *-t,* and the 2nd person plural verbal marker is *-ðix* (< *tik* ?).

2.14 Personal pronouns in Etruscan

Among Etruscan pronouns, only the 1st person singular is known (Table 12). It displays an alternation between a nominative form *mi* and a non-subject form *mini.* The presence of a root *m-* plus a non-subject suffix *-n* is a strong indication that Etruscan belongs to the Eurasiatic macrofamily. According to Greenberg, it might even indicate that Etruscan is in fact Indo-European, although one might note that the particular form of the Etruscan alternation is more like that canonical in Uralic, e.g., Finnish *mä ~ minä* "I" opposed to *minä* "me."

Table 12. 1st person *m-* in Etruscan personal pronouns

Language	**1st sg.**		**1st pl.**	**2nd sg.**	**2nd pl.**
Etruscan	*mi*	*mini*	**?**	**?**	**?**

Conclusion

As in Indo-European, a 1st person pronoun *m-* is attested in all member languages of the Uralic-Yukaghir, Altaic, Chukchi-Kamchadal and Eskimo-Aleut families (plus Etruscan). Moreover, all these six language families attest to a root alternation between the nominative and non-subject forms, together with a suffix *-n* in the non-subject form. In Nippo-Koreo-Ainu and Gilyak, its presence is less certain, although the Ryukyuan *baa ~ banu* alternation, though deprived of *m-,* closely resembles the Altaic pattern, while Gilyak dual *megi* "the two of us" and plural *mer* "we," with their clearly Eurasiatic dual *-ki* and plural *-ri* suffixes, also seem pretty secure.

Thus, it is not only a common word root that these language families share with the 1st person *m-,* but a complex grammatical pattern involving a frequent root suppletion in the nominative and a suffix *-n* in other cases, certainly having originated as a genitive. Is it possible to imagine that such a complex pattern might have arisen six times by chance (if we leave aside the less obvious Nippo-Koreo-Ainu and Gilyak families)? It seems highly unlikely. As to the 2nd person pronoun *t-*, it is somewhat

less widely represented in Eurasiatic languages. Nevertheless, it is preserved in all languages of Indo-European, Uralic-Yukaghir, the Mongolic branch of Altaic, Chukchi-Kamchadal, and Eskimo-Aleut. It also left strong traces in the Tungusic branch of Altaic and perhaps in Gilyak, while it completely lacks in the Turkic branch of Altaic and the whole Nippo-Koreo-Ainu family – as well as in Etruscan because of our poor knowledge of this language. Once we add this 2nd person *t-*, plus the different suffixes briefly mentioned above, it seems that there is no more room for doubt – except for Nippo-Koreo-Ainu and perhaps Gilyak: the 1st and 2nd person pronouns we have studied (and, consequently, the languages to which they belong) must descend from a common origin.

Among the different suppletive nominative roots of 1st person pronoun, the close resemblance between Indo-European *eg^{h}om* "I" and Chukchi-Kamchadal *(e-)ɣəm* "I" is striking. Together with the fact that these two families are the only ones to attest *tu* for the 2nd person, it might suggest a special historical relationship between Indo-European and Chukchi-Kamchadal.

A geographic observation may find its proper place here. The language families concerned are not randomly distributed over the Americas, Australia, and Africa. Rather, they constitute a geographic continuum, exactly as if they had resulted from an expansion over northern Eurasia. This expansion, we might add, must have started from a homeland which is most probably to be sought somewhere near the northern Pacific coast of Asia, according to the law of maximal diversity area. This law, long-known in biology, was discovered in linguistics by Greenberg (1972) in his controversy with Guthrie (1970) about the Bantu homeland.

The age of the Eurasiatic macrofamily is impossible to assess precisely, except that it is of course older than Indo-European, so that the 1st person pronoun *m-* was preserved in the overwhelming majority of the descendant Eurasiatic languages through at least some 10,000 years or more. It makes definitely clear that (1) calculations based on an average word replacement rate are highly uncertain, and that (2) particular words may be preserved through almost indefinite timespans. Regarding our claim that the globally distributed *papa, tata, mama, nana* or *kaka* kinship terms have been inherited from an original Proto-Sapiens language, a consequence of these two conclusions is to invalidate the objections levelled at their common ancestry on the ground of word replacement rates.

3. Kinship terms and the origin of 1st and 2nd person pronouns

Eurasiatic pronouns *m-* and *t-* are built from consonants that are also the root consonants of the Proto-Sapiens kinship terms *mama* and *tata.* It appears that the case is far from being isolated in the world's languages. A cursory count through Ruhlen's (1994b) global survey of 1st and 2nd person pronouns reveals that about three quar-

ters of them are based on one of the consonants *t, k, m, n, j, ŋ*, and *s*. The first five are the root consonants of the globally-distributed kinship terms *tata, kaka, mama, nana* and *jaja*. As to the last two, namely *ŋ* and *s,* they are also the base of widely, though not fully globally distributed kinship terms. In particular, the *s* is the root consonant of the widespread *ise* "father, father's brother, brother," present at least in the Eurasiatic, Amerind and Niger-Congo language families. No other consonantal type is widespread among 1st and 2nd person pronouns.

Viewed from the kinship terms' side, only the *p ~ b* oral labial stop of the highly widespread *papa ~ baba* is extremely rare among the world's 1st and 2nd person pronouns. (Among the few exceptions are the suppletive *b-* forms in Altaic 1st person pronouns, which are evidently derived from an original *m-* root.) The near absence of *p ~ b* among the thousands of these pronouns in the world's languages, while almost no language lacks an oral labial stop, is typologically highly unexpected. Given that kinship terms and personal pronouns are the two most conservative word categories, and consequently are in general extremely ancient in each language family, two questions arise. The first question is whether pronouns could be traced back to Proto-Sapiens through a global comparison, as was already done for globally-distributed kinship terms. Answering this question would be much more difficult than for kinship terms, for at first glance personal pronouns in the world's languages look like a horrible mess of *m* "I" and *m* "thou," *n* "thou" and *n* "we," and so on. We will thus leave the question open.

Answering the second question will perhaps help to answer the first one – one day. This second question is whether there could or must be a historical link between the category of personal pronouns and that of kinship terms. And our answer is yes.

Nursery kinship terms, as we argued elsewhere (Bancel & Matthey de l'Etang 2004, 2005; Matthey de l'Etang, Bancel & Bengtson 2006), are definitely of Proto-Sapiens ancestry – because of their ubiquity and the impossibility that they had resulted from convergent innovations. Contrary to a widespread belief, such innovations are unattested in the low-level language families with a written history (Matthey de l'Etang & Bancel this volume). Nursery kinship terms must even be much more ancient than Proto-Sapiens, and certainly played a major role in the apparition of Proto-Human – language with a simple, not double, articulation, i.e., with words built from phonemes but without sentences built from words (Bancel, Matthey de l'Etang & Ruhlen 2006). The first phonetically articulated words, uttered by mouths and tongues that had not been designed for speech by evolution, must have been built from the simplest consonants cast into the simplest syllable structures. They also must have carried highly useful meanings derived from pre-language social functions. They also must have been easy to transmit from generation to generation through mouths, brains and ears lacking specialisation for language, so that this invention did not get lost. All these conditions are fullfilled by nursery kinship terms

(Lieberman *et al.* 1972; Lieberman 1992; Bancel & Matthey de l'Etang 2005), and by them only.

Personal pronouns, in turn, must be of much more recent origin, for at least two compulsory reasons. The first reason is that, before the apparition of articulated sentences, they were simply useless. As long as articulate language merely consisted in isolated, holophrastic words, or, in a later stage, in parasyntactic sequences comprising a small number of juxtaposed words, to design and use specific words to refer to the speaker and the hearer in relation with verbs and actions was not only aimless, it was plainly unconceivable. The development of personal pronouns from the already existing vocabulary may only have taken place in a much later stage, once speakers had begun to gather words into sentences and the most frequent combinations had begun to crystallize into syntax.

The second reason is that personal pronouns are conceptually the most difficult words – much more than *quantum* or *phenomenology.* Not that we mean the *ego* is such an unfathomable mystery that only an Austrian genius could have invented it in the beginning of the XXth century (while the *tu* still keeps waiting). Rather, the special difficulty of personal pronouns resides in their WAY OF MEANING. In all languages, they are not the only words whose meaning depends on the speaker – after all, the English words *here* and *now* respectively mean "(close to) the place where I am" and "at the moment when I am speaking." But personal pronouns are the only words whose reference EXCLUSIVELY CONSISTS OF the speaker (for the 1st person) and the hearer (for the 2nd), so that their reference automatically switches as the speech turn passes in the course of conversation.

Does this specificity really make pronouns difficult words? The fact that children learn to use them in a relatively late stage of language acquisition (around 3 years), after they have learned most of the other basic syntactic structures, indicates that such is indeed the case. Even older children of 4 to 5 years, apparently mastering the specific referential mode of personal pronouns, are easy to confuse when an adult playfully acts as if pronouns were normal words, i.e., as if he himself was "an I" and the child was "a you," and "corrects" the child each time he uses the pronouns the right way.

For other words, the reference remains the same, whoever is speaking. If I tell you about the colorless dog furiously sleeping on the green carpet near the door, and you answer me this sleeping dog on the carpet is a degenerate subject, in both sentences – yours and mine – the words *dog* and *carpet* refer to the same beings, and the word *sleeping* refers to the same action. Most other words, and in particular all common and proper nouns, function the same way. And among the thousands of common nouns, it is also true of *quantum* and *phenomenology.* If the semantic content of these two words is certainly difficult, their semantic container is strictly trivial.

Kinship terms, in turn, refer in a specific way. In English, for instance, *father* is a common noun, but *dad* is and is not. If I ask you about *"your dad,"* it is a common

noun, but if I inform you that *"Dad is out fishing,"* it is a proper noun referring to a single person – a specificity aptly rendered, in the usual typography, by the use of an initial capital. You understand that *Dad* is in fact my father, and if you reply *"Oh! Dad went out early, they must have gone together,"* I understand in turn that you are speaking of your own father. Thus, kinship terms intrinsically share properties with all three nominal categories of proper nouns, common nouns, and pronouns. Like proper nouns, they can refer to a specific individual. Like common nouns, they also can refer to a class of beings, defined by common properties of these beings. And, like pronouns, they can switch reference from a given individual to another as the speech turn passes.

Now, at the global level, we find the following picture. All languages have kinship terms, common nouns, proper nouns, and personal pronouns. Among the four categories, kinship terms certainly are the oldest ones. Common and proper nouns are certainly very ancient, and have appeared in the course of the Proto-Human language evolution as natural developments from the primeval kinship terms. Finally, personal pronouns may not have appeared before a much later stage, when language verged on double articulation – i.e., on language in the modern sense, with sentences made of syntactically articulated words, as are all known languages. Facing this general picture, it appears that kinship terms and 1st and 2nd person pronouns share most of their root consonants at the global level. A straightforward, commonsense explanation would certainly be that both categories are based on the most frequent consonant types, and that their convergence is thus purely coincidental. This explanation, however, faces two problems. The first one is that both kinship terms and pronouns are extremely resistant to change, as we have seen with Indo-European and Eurasiatic pronouns, and thus have been, in most cases, inherited from ancient ancestral languages – a fact in and by itself hardly compatible with the idea of a random distribution of their constituent phonemes. The second one is the near absence, among the world's 1st and 2nd person pronouns, of oral labial stops *p* ~ *b.* A random distribution of pronouns could hardly have missed to use these very basic sounds. Conversely, the absence of *p* ~ *b,* if it could be regarded as a probabilistic artifact in a single language, would be very unlikely if pronouns originated in several distinct ancestor languages.

Another explanation is that there is indeed a relation between the two categories. The Proto-Human language certainly made do without pronouns for a long time. Even a fully (i.e., syntactically) articulate language can function without personal pronouns, for instance using either common or proper nouns to refer to the speaker and the hearer, like in sentences such as *"One-eyed hunter kill big mammoth at black cliffs for clan"* or *"Sweet Rob bring newspaper from office for Darling Janie."* However, all known languages have personal pronouns. At some point in the evolution of Proto-Human to modern languages, speakers must have evolved personal pronouns from already existing words. Which words? These precursors of pronouns must have been nominals, which in the preceding stage were the most used words designating the speaker and the hearer – words of the kind of *one-eyed hunter, clan, Rob,* or *Janie.*

However, if either common or proper nouns had given rise to pronouns, the global comparative picture of present-day pronouns would be very difficult to explain. If all modern pronouns shared a common origin, and had descended from a subset of common and/or proper nouns in a single ancestor language, how could one explain that it seems impossible to assign any of the modern pronominal root consonants to a global origin? It seems at odds with the exceptional preservation of pronouns we have dealt with above. And if present-day pronouns descended from a subset of proper or common nouns in several different original languages, how could one explain that they converge so massively towards a handful of root consonants, whatever the language family they belong to?

It happens that a third nominal category may serve to the same purpose – kinship terms. And they are by far the most likely category to have given rise to personal pronouns. In prehistoric conditions, all humans lived in small groups of kindred individuals – as is the case of all historically known hunting-gathering peoples as well as of our closest cousins, chimpanzees and bonobos. In the stage of Proto-Human language that preceded the apparition of pronouns, kinship terms such as *mama, tata, nana* or *kaka* may have been the most frequent way to address people, so that one of them might easily have given rise to a 2nd person pronoun. It may seem less straightforward for the 1st person pronoun, since by definition there is no kinship term referring to oneself. However, just like for 2nd person, the 1st person pronoun must have emerged from an earlier nominal used by the speaker to refer to himself, and no other nominal category possesses such a word. It is perfectly conceivable that, in the stage before the emergence of personal pronouns, speakers referred to themselves by the kinship term used towards them by the addressee. In modern languages with personal pronouns, such practice may seem weird, but is occasionally used when speaking with children who do not master the use of personal pronouns, as in *"Mum wants Sonny to eat up those peas."* It would explain why so many present-day languages have pronouns built from the same consonants as the most widespread kinship terms, and in our opinion it is the only way to explain this striking convergence consistently.

However, several crucial questions remain to be answered. How was performed the transition from the masculine~feminine meanings of kinship terms to the 1st ~ 2nd person meanings of pronouns? Why, contrary to kinship terms, do the pronominal root consonants of present-day languages not display any evident semantic convergence? And why are *p* ~ *b* consonants almost lacking as pronominal roots, while *papa* ~ *baba* kinship terms are extremely widespread?

To the first question, we do not have the faintest beginning of an answer. With regard to the two others, we only may make speculative suggestions. The apparent lack of semantic convergence for pronominal root consonants at the world level might be due to the fact that, at the time of the initial dispersal of our Sapiens ancestors, their language only had begun to develop the syntactic articulation – which gave them a fantastic technical impulsion, as suggested for the first time

by Ruhlen (1997). The category of pronouns was not yet definitely fixed, so that not exactly the same set of kinship terms crystallized into pronouns in the different language families.

Regarding the lack of *p* ~ *b* consonants in 1st and 2nd person pronouns, one of our present puzzles regarding global kinship terms might perhaps give us a solution. Apparently, different Proto-Sapiens etymologies referred to the same kinship classes. We are able to reach a meaning like "male elder on the father's side" for both *papa* and *tata* in Proto-Sapiens, on the basis of more specialized meanings in modern languages, such as "father," "father's brother," "brother" and "grandfather." Taken together, these four meanings represent more than 70% of languages where *papa* (or *tata*) forms are attested, while the remaining 30% of languages use their *papa* or *tata* forms with many different meanings with no statistical significance (Bancel & Matthey de l'Etang 2004; Matthey de l'Etang & Bancel 2005). Thus, *papa* and *tata* appear as synonymous, and so do the three globally-distributed *mama, nana* and *jaja* "female elder on the mother's side." But, even though the difference between them escapes reconstruction at present, they certainly were not true synonyms – otherwise they would not have all survived. The widespread distinction in modern languages between "my father" and "your father," and between "my mother" and "your mother," sometimes with affixes but also often with two distinct terms, could be a trace of this putative original difference. Whatever this difference may have been, *papa* may have had a specificity that led to exclude it when referring to both the speaker and the hearer, so that it would not have given rise to a 1st nor 2nd person pronoun. This would explain the strange scarcity of oral labial consonants among the variety of pronominal roots in modern languages.

Finally, our conjecture may be condensed as follows:

Given that

(1) kinship terms must be of the highest antiquity in the development of articulate language, while pronouns may have appeared only at a much later stage;

(2) pronouns must have emerged from nominals frequently used to refer to the speaker and the hearer; and

(3) the pronominal root consonants at the global level are a subset of those of kinship terms,

It is difficult to see how personal pronouns could have emerged otherwise than from simplified kinship terms formerly used to represent the speaker and the hearer in discourse.

Conclusion

We have shown that resistant words such as the Indo-European and Eurasiatic pronouns may last for almost indefinite timespans. Their persistence, paralleled in the overwhelming majority of language families, gives credibility to the possibility that kinship terms might have lasted a hundred millennia, or more. We have then put forward the conjecture that there could be a relationship between kinship terms and 1st and 2nd person pronouns. Even if this conjecture should remain unproven and unrefuted for a while, it shows that we have linguistic means to investigate the evolution of human language from its earliest stages.

References

Bancel, P.J. & Matthey de l'Etang, A. 2002. "Tracing the Ancestral Kinship System: The Global Etymon KAKA. Part I: A Linguistic Study." *Mother Tongue* **7:** 209–243.

Bancel, P.J. & Matthey de l'Etang, A. 2004. "Kin Terms *(P)APA, (T)ATA, (M)AMA, (K)AKA* and the Origin of Articulate Language." Paper presented at the Language Origins Society 18th Annual Meeting, Oxford (Eng.), 9–10 July 2004. Appeared in 2006 in a special issue of *Cognitive Systems*, Groningen.

Bancel, P.J. & Matthey de l'Etang, A. 2005. "Kin Tongue. A Study of Kin Nursery Terms in Relation to Language Acquisition, with a Historical and Evolutionary Perspective." *Mother Tongue* **9:** 171–190.

Bancel, P.J., Matthey de l'Etang, A. & Ruhlen, M. 2006. "The Proto-Human Words PAPA and MAMA and the Origin of Articulate Language." Paper presented at the 105th American Anthropological Association Annual Meeting, to appear in the proceedings of the Per Hage Memorial Session.

Bengtson, J.D. & Ruhlen, M. 1994. "Global Etymologies." In Ruhlen 1994b: 277–336.

Benveniste, E. 1946. "Structure des relations de personne dans le verbe." *Bulletin de la Société de Linguistique de Paris* **43**–1. Repr. in *Problèmes de linguistique générale* 1: 225–236, Paris: Gallimard.

Bomhard, A.R. 2007. *Reconstructing Proto-Nostratic. Comparative Phonology, Morphology, and Vocabulary*. Charleston: Signum.

Dolgopolsky, A. 1964. "Gipoteza drevnejšego rodstva jazykovyx semej severnoj Evrazii s verojatnostnoj točki zrenija." *Voprosy Jazykoznanija* **2:** 53–63. [Translated into English as "A Probabilistic Hypothesis Concerning the Oldest Relationships among Language Families of Northern Eurasia." *Typology, Relationship and Time*, Ed. by V.V. Shevoroshkin and T.L. Markey, 27–50. Ann Arbor: Karoma. 1986.]

Greenberg, J.H. 1972. "Linguistic Evidence Regarding Bantu Origins." *The Journal of African History* **13**–2: 189–216.

Greenberg, J.H. 2000. *Indo-European and Its Closest Relatives. The Eurasiatic Language Family.* Vol. I: *Grammar*. Stanford, Calif.: Stanford University Press.

Greenberg, J.H. 2002. Indo-European and Its Closest Relatives. *The Eurasiatic Language Family.* Vol. II: *Lexicon.* Stanford, Calif.: Stanford University Press.

Guthrie, Malcolm. 1970. "Contribution from Comparative Bantu Studies to the Prehistory of Africa." *Language and History in Africa*, Ed. by D.E. Dalby, 20–49. London: Frank Cass.

Lieberman, P. 1992. "On the Evolution of Human Language." *The Evolution of Human Languages,* Ed. by J.A. Hawkins & M. Gell-Mann, 21–47. Redwood City, Calif.: Addison-Wesley.

Lieberman, P., Crelin, E.S. & Klatt, D. 1972. "Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man and Chimpanzee." *American Anthropologist* **84:** 287-307.

Matthey de l'Etang, A. & Bancel, P.J. 2005. "The Global Distribution of (P)APA and (T)ATA and their Original Meaning." *Mother Tongue* **9:** 133–169.

Matthey de l'Etang, A. & Bancel, P.J. 2008. "The Age of MAMA and PAPA." This volume.

Matthey de l'Etang, A., Bancel, P.J. & Bengtson, J.D. 2006. "The Proto-Sapiens Words PAPA, MAMA and KAKA and the Origin of Kinship Systems." Paper presented at the 105th American Anthropological Association Annual Meeting, to appear in the proceedings of the Per Hage Memorial Session.

Pagel, M. 2000. "Maximum Likelihood Models for Glottochronology and for Reconstructing Linguistic Phylogenies." C. Renfrew, A. McMahon & Larry Trask (Eds), *Time Depth in Historical Linguistics*, Cambridge (Eng.): McDonald Institute for Archaeological Research I: 189–207.

Rosenfelder, M. No date. "Numerals in 5,000 languages." http://www.zompist.com/numbers.shtml.

Ruhlen, M. 1991. *A Guide to the World's Languages,* vol. I: *Classification*. Stanford, Calif.: Stanford University Press.

Ruhlen, M. 1994a. *The Origin of Language. Tracing the Evolution of the Mother Tongue*. New York: John Wiley.

Ruhlen, M. 1994b. *On the Origin of Languages. Studies in Linguistic Taxonomy*. Stanford, Calif.: Stanford University Press.

Ruhlen, M. 1997. Public conference given in Lyons for the appearance of the French translation of Ruhlen (1994a).

Ruhlen M. 2007. Post-Script to the 2nd edition of *L'Origine des langues. Sur les traces de la langue mère* [Fr. tr. of Ruhlen (1994*a*)]. Paris: Gallimard.

Starostin, S.A., Dybo, A.V. & Mudrak, O.A. 2003. *Etymological Dictionary of the Altaic Languages.* Leiden: Brill.

Starostin, S.A., Starostin, G.S., Nikolayev, S.F., Dybo, A.V., Mudrak, O.A. & Golovastikov, A. No date. "Nostratic Etymological Database (including Indo-European, Altaic, Uralic, Kartvelian, Dravidian, Chukchee-Kamtchatkan, and Eskimo Etymological Databases). Online at: http://starling.rinet.ru/index2.php?lan=en.

Strand, R. 1997–2008. *Nuristân. Hidden Land of the Hindu-Kush.* http://users.sedona.net/~strand/.

Swadesh, M. 1952. "Lexico-Statistic Dating of Prehistoric Ethnic Contacts, with Special Reference to North American Indians and Eskimos." *Proceedings of the American Philosophical Society* **96** 4: 452–463.

Swadesh, M. 1954. "Time Depths of American Linguistic Groupings." *American Anthropologist* **56** 3: 361–377.

Thomason, S.G. & Everett, D.L. 2005. "Pronoun Borrowing." *Proceedings of the Berkeley Linguistics Society 27th Annual Meeting* 301–315. Berkeley: Berkeley Linguistics Society. Online version http://www-personal.umich.edu/~thomason/papers/pronborr.pdf.