

Journal of Universal Language 8
September 2007, 39-59

The Kaingang (Brazil) Seem Linguistically Related to Oceanic Populations*

Vladimir Pericliev
Bulgarian Academy of Sciences

Abstract

The Southern-Ge Kaingang people, comprising the Xokleng and the Kaingang, presently reside in the states of Santa Catarina, São Paulo, Paraná, and Rio Grande do Sul in south-eastern Brazil, and speak two closely related languages, forming the Kaingang language family. No external relationship of the Kaingang people, or their language, with peoples or languages outside of Ge, or Amerind(ian)

* Part of this work was conducted in 2002 under EC Marie Curie grant HPMF-CT-2001-01409 at the Max Planck Institute for Evolutionary Anthropology, and thanks go to my hosts there, geneticist M. Stoneking and linguist B. Comrie for their hospitality. Linguist U. Wiesemann kindly supplied some information on Kaingang and Xokleng and anthropologists D. White and D. Read ensured access to the original G.P. Murdock files on Xokleng kinship. I am however especially indebted to physicist Andrey Karshikoff (Karolinska) and to linguist Bengt Sigurd (Lund University) for both fruitful discussions and encouragement. Support is also gratefully acknowledged from grant MI-1511 with the Bulgarian Ministry of Education and Science.

more generally, has so far been suggested. In a previous paper in this journal (Pericliev 2006), I showed that statistically significant lexical similarities exist between Xokleng and several Oceanic languages, and based on further linguistic and population genetic data, tentatively put forward the hypothesis concerning the existence of a historical relationship between the Brazilian language and Austronesian. In this paper, I will show significant kinship similarities between Xokleng and Austronesian, as well as sketch some grammatical similarities. Then, including also the Kaingang language into the comparison, more than thirty putative cognate words will be given alongside with sound correspondences between the Kaingang and the Polynesian language families seeming to corroborate the idea of a link between the Brazilian languages on the Atlantic and the Oceanic languages to a considerable extent.

Keywords: Austronesian-Kaingang relationship, Oceanic and South American contacts, language prehistory and classification, languages and migrations

1. Introduction

Contemporary linguistics disposes with generally reliable sources of language classification (Gordon 2005, Ruhlen 1987). The problem of language classification, however, is not closed. Thus, on the one hand, there still exist unclassified languages and, on the other hand, some of the classifications proposed are controversial (cf. esp. the internal groupings within well-understood language families, or so-called “macro-families” lumping together other families). The choice of which languages exactly to compare for potential link is computationally difficult, and beyond human reach, in view of the immense number of logically possible alternatives resulting from combinations of the 5-6 thousand languages known to exist today.

Thus, in early 2002, I implemented a computer program that was intended to aid the linguist in the task of discovering statistically

significant (i.e., non-chance) similarities between languages. Finding non-random resemblances between languages believed to belong to distinct language families then could be interpreted as a presence of a “historical relationship” between these languages. Such a historical relationship between languages, in principle, may either be due to genetic reasons (i.e., owing to common ancestry) or due to contact and borrowing.

The results from running the program on a kinship terminology database most unexpectedly revealed non-chance similarities between the geographically distant Brazilian language Xokleng (with classification according to *Ethnologue* as Macro-Ge/Ge-Kaingang/Kaingang/Northern) and several Austronesian languages. The languages were too distant to assume contact, and no such contact is historically known. This experiment will be described in Section 2 of the present article. This interesting computationally-generated conjecture was further pursued and some diagnostic structural features were found to exist between Xokleng and Austronesian, which will be sketched in Section 3. Computational comparisons of wordlists, presented in a previous article in this journal (Pericliev 2006), were also statistically significant and added further credibility to the conjecture, alongside with other linguistic and population genetic data, but as I argued there additional evidence is needed to get more definitive answers. Section 4 will present over thirty putative cognates, with sound correspondences, between the Kaingang language family (Xokleng and Kaingang) and the Polynesian language family, exemplified by Māori, and Hawaiian. The evidence seems to corroborate my hypothesis to a considerable extent. Section 5 concludes, noting that further research is needed in this direction.

2. Kinship Similarities

A computer program was implemented accepting as input languages, described in terms of feature-values, and disposing of information as to the genetic affiliation of each, compares all pairs of languages, belonging to *different* language families. For each pair, the program computes the statistical significance of the occurrence of their common feature-values, using to this end the permutation method, as described in Valdés-Pérez and Pericliev (1999). The program discards as uninteresting/chance any language pair whose feature-value associations (=similarities) are not statistically significant. If, in contrast, the feature-value overlap of two languages belonging to two different language families (according to current knowledge) turns out to be statistically significant (=non-chance), the program outputs that pair of languages as probably “historically related”.

The program used as data the G. P. Murdock (Murdock 1970) database of kinship semantic patterns. G. P. Murdock describes the terminological classification system of 566 languages from 194 of the 200 cultural provinces that he had previously isolated. This data set is the most representative compilation of kinship terminologies to date. The data includes virtually all systems published for Africa and aboriginal North and South America, and is only slightly less exhaustive for Eurasia and Oceania. Moreover, the data set is based on files of over a thousand complete systems; the published data set includes only those systems which differ from the remaining systems within the same sampled province in order to evade duplication.

The Murdock data set focuses on eight sets of kin (“features” in our sense): *grandparents* (GrPa), *grandchildren* (GrCh), *uncles* (PaBr), *aunts* (PaSi), *nephews and nieces (male speaker, ms)* (SbCh), *siblings* (Sibl), *cross-cousins* (CrCo), and *siblings-in-law* (Sb-Inl). Every type of kin is described in terms of “kin term patterns”

(“features-values” in our sense), showing the number of kin terms used for that kin as well as their range of reference.

We may consider some examples of kin term patterns. For “grandparents”, for instance, Murdock gives 20 patterns in all, the first six of which (preserving for reference the original notation) are as follows:

- (1) *Bisexual Pattern*. Two Terms, Distinguished by Sex, which can be Glossed as “grandfather” and “grandmother”.
 - a. Variant of A, with separate terms for GrFa (ms), GrFa (woman speaker, ws), and GrMo.
- Merging Pattern*. A Single Undifferentiated Term, which can be Glossed as “Grandparent”.
 - a. Variant of B with separate term for MoMo only.
 - b. Variant of B with separate term for MoMo (ws) only.
 - c. Variant of B with separate terms for FaFa (ms) and FaFa (ws).

For the eight sets of kin Murdock describes, he uses 192 patterns in all, distributed as follows: GrP(20), GrCh(20), PaBr(13), PaSi(14), SbCh(26), Sibl(43), CrCo(18) and Sb-Inl(38).

For our purposes we needed to associate each language in Murdock’s sample with the language family to which this language belongs. To this end, *Ethnologue* (Gordon 2005), a standard and constantly updated reference on world languages and language families, was used.

The results from running the program may be summarised as follows. Three language pairs turned out to have similarities which are “highly significant” ($p < 0.01$), or what is the same we have assurance of at least 99% that these similarities are non-chance. Four language pairs turned out to have similarities that are “significant” ($p < 0,05$), or for which we have assurance of at least 95% that their similarities are non-chance. Below I list these

language pairs, together with the language families they belong to according to *Ethnologue*. The number of overlapping kinship semantic patterns is also given (the particular content of these patterns can be checked and their explanation found in Murdock 1970).

- (2) a. Highly Significant Similarities between Language Pairs ($p < 0.01$)
 1. Xokleng (Macro-Ge) – Ami (Austronesian), seven common patterns
 2. Xokleng (Macro-Ge) – Trukese (Austronesian), six common patterns
 3. Xokleng (Macro-Ge) – Ulithian (Austronesian), six common patterns
- b. Significant Similarities between Language Pairs ($p < 0.01$)
 4. Rwala-Bedouin (Afro-Asiatic) – Anatolian-Turkish (Altaic), seven common patterns
 5. Iban (Austronesian) – Khmer (Cambodia) (Austro-Asiatic), seven common patterns
 6. Maria-Gond (Dravidian) – Baiga (Indo-European), seven common patterns
 7. Icelandic (Indo-European) – Egyptian (Afro-Asiatic), eight common patterns

Many borderline significant similarities between pairs of languages emerged, but I ignore them inasmuch as I am interested here only in commonalities that are very unlikely to have occurred simply by chance.

Cases 4-7 above are suggestive of some kind of historical relationship between the language pairs. For example, as regards the Indo-European language Baiga and the Dravidian language Maria-Gond (case 6), one may speculate that perhaps contact and borrowing are involved inasmuch as both languages are spoken in the Indian region Madhya Pradesh. The Indo-European language Icelandic and the Afro-Asiatic language Egyptian (case 7) share all 8 compared patterns, and are too distant in space to assume borrowing; and indeed, genetic relationship between these two families is sometimes posited in the literature. Austronesian and Austro-Asiatic (cf. case 5) are also believed by some linguists to be genetically related under the Austric macro-family. Both most reliable and most interesting, however, are undoubtedly cases 1-3 and below I turn to their discussion.

Cases 1-3 above suggest some historical relationship between the Macro-Ge language Xokleng, spoken in the eastern, Atlantic, part of Brazil and three Austronesian languages, Ulithian, Trukese, and Ami, spoken in the Pacific.

Below is an explanation of Xokleng's kinship semantic patterns. For ease of reference, further on I shall use their abbreviations (note that patterns denoted by initial letters in the alphabet, A, B, C are used for common and those denoted by final letters are used for rare patterns; e.g. Sb-Inl=A means that the pattern A for siblings-in-law is the most common one in the database, Sb-Inl=B that the pattern B for siblings-in-law is the next common, and so on).

- (3) a. GrPa=L. "Null Pattern", in which special terms are lacking for grandparents, who are called by the same terms for parents.
- b. GrCh=K. "Null Pattern", in which special terms are lacking for grandchildren, who are called by the same term or terms that the speaker applies to his own children.
- c. PaBr=E. "Generation Pattern", in which special terms are

lacking for both paternal and maternal uncles, who are terminologically equated with father.

- d. PaSi=D. “Generation Pattern”, in which special terms are lacking for both paternal and maternal aunts, who are terminologically equated with mother. Analogous to Pattern E for uncles.
- e. SbCh=F. “Sex-Differentiated Lineal Pattern”, in which there are two special terms, differentiated by sex, which can be glossed by as “nephew” and “niece”.
- f. Sibl=C. “Yoruban Pattern”, in which there are two terms, distinguished by relative age, which can be glossed as “elder sibling” and “younger sibling”.
- g. CrCo=A. “Hawaiian Pattern”, in which special cousin terms are lacking, both cross and parallel cousins called by the terms for siblings.
- h. Sb-Inl=A. “Merging Pattern”, in which there is a single undifferentiated term, which can be glossed as “sibling-in-law”.

Xokleng exhibits seven identical patterns with Ami, out of eight types of patterns compared. These are the patterns GrCh=K, PaBr=E, PaSi=D, SbCh=F, Sibl=C, CrCo=A, and Sb-Inl=A. The language shows six matches with Trukese (viz. GrPa=L, GrCh=K, PaBr=E, PaSi=D, SbCh=F, Sb-Inl=A) and six matches with Ulithian (viz. GrPa=L, GrCh=K, PaBr=E, PaSi=D, SbCh=F, CrCo=A). Besides these highly statistically significant overlaps, we may mention some other Austronesian languages that notably resemble Xokleng, e.g. Malayan and Rotuman with five overlaps, Samoan, Merina, Kapingamarangi, Māori with four overlaps, etc.

Xokleng thus seems to follow an Austronesian type of kinship semantic patterning (in his original files, G. P. Murdock marks the structural type of Xokleng’s terminology as “Normal Hawaiian”, which is characterised by a generational-terminology structure for

the parental generation, and in which there is no distinction between siblings and cousins, all called by the term for sibling). Table 1 summarizes the distribution of Xokleng's patterns in Austronesian languages on the one hand, and in all remaining languages (without the Macro-Ge ones) on the other.

Table 1. Comparison of the distribution of the patterns of Xokleng in Austronesian and in the remaining languages.
(Total number of: Austronesian languages=84,
All remaining languages without Macro-Ge=473)

| Patterns of Xokleng | Austronesian lgs with same pattern | | All remaining lgs with same pattern | | |
|---------------------------------|------------------------------------|-------------|-------------------------------------|---------|----------------------------------|
| | Number | Average | Number | Average | Distributed in No of lg families |
| GrPa=L Null | 4 | 5.0 | 2 | 0.4 | 2 |
| GrCh=K Null | 9 | 11.2 | 10 | 2.0 | 7 |
| PaBr=E Generation | 18 | 22.5 | 9 | 1.8 | 7 |
| PaSi=D Generation | 30 | 37.5 | 33 | 6.7 | 11 |
| SbCh=F Sex-Differ. Lineal | 18 | 22.5 | 34 | 6.9 | 17 |
| Sibl=C Yoruban | 15 | 18.8 | 32 | 6.6 | 15 |
| CrCo=A Hawaiian | 38 | 47.5 | 136 | 28.0 | 35 |
| Sb-Inl=A Merging | 19 | 23.8 | 52 | 10.7 | 17 |

As seen on Table 1, some of the rarer patterns of Xokleng (viz. GrPa=L, PaBr=E) are more frequent in Austronesian than in all remaining languages, even though the Austronesian languages in the examined database are only 84, while all remaining languages (without the Macro-Ge ones) are 473. Other rare patterns of Xokleng (viz. GrCh=K, PaSi=D) are comparably frequent in Austronesian and in all remaining languages (also bearing in mind the smaller number of Austronesian as compared to all remaining languages). For all eight investigated types of patterns, Xokleng shares a pattern that is, on the average, more common for Austronesian than for the class of all remaining languages (compare column 3 with column 5; the larger value of the two columns is given in bold). The patterns that are more frequent in absolute terms in non-Austronesian languages than in Austronesian seem to be distributed more or less randomly among a large number of language families.

In sum, in pair-wise comparisons, the Brazilian language Xokleng exhibits striking similarities in its kinship semantic patterns to each of the languages Ami, Ulithian, and Trukese, all belonging to the Austronesian language family. In each case, these similarities are statistically very significant (i.e., highly unlikely to have occurred simply by chance); the probability that the JOINT occurrence of all three events is only chance is practically nil. Besides these commonalities with the mentioned languages, Xokleng resembles significantly other Austronesian languages, showing a very Austronesian type of kinship system, labelled “Normal Hawaiian” by the famous anthropologist G. P. Murdock.

3. Phonological and Grammatical Similarities

Some features of the language of the Xokleng (Henry 1935, 1948; Urban 1985, 1986) were tested against a set of typically Austronesian features that have been proposed (Klamer 2002) as a heuristic for suggesting the affiliation of a language (i.e. whether it is Austronesian or not). Xokleng turned out to be remarkably Austronesian-like, sharing Austronesian properties of both phonology and grammar.

Like Austronesian languages, and especially their Oceanic branch, which are known to have lost the voicing contrast in obstruents and to have developed prenasalized consonants in opposition to plain consonants, Xokleng (as described in Henry 1935, 1948) also does not have plain voiced obstruents, but contrasts plain voiceless with prenasalized voiced consonants.

Many Austronesian languages prefer roots of CVCV type, and so does Xokleng. According to Henry (1948: 196) the CVCV pattern is prevalent and amounts to 35% of all root patterns in Xokleng (CV patterns being 14%, CVC 13%, CVCVC 12%, CVCCV 12%, CCV 5%, and CVCVCV 5%). As seen from these numbers, other typically Austronesian features of Xokleng are the "dropping" of final consonants (word-final consonants being present in only 25% of the patterns) and a dispreference for consonantal clusters (occurring in only 17% of the words). Besides, the possible final consonants and consonantal clusters are subjected to further restrictions we need not discuss here.

Similarly to many Austronesian languages, Xokleng forms emotional expressions by Verb + body part noun, in which the Experiencer of the emotion is the Possessor of the body part. E.g. Xokleng's expression for "I am angry" literally means "My heart splits (in several places)" (Henry 1935: 213).

Like many Austronesian languages, Xokleng's numerals seem to

behave like verbs in that they act like verbal predicates and take the same predicating particles as the verbs.

Also like other Austronesian languages mainly in Eastern Indonesia, but also scattered elsewhere in Western and Eastern Austronesian languages (e.g., Malagasy, Manobo, Hawaiian, Batak, etc.) Xokleng (cf. Urban 1986) employs parallelism in narratives, myths, poems and songs, a verbal art form in which semantically synonymous pairs/triples etc. of words and phrases are used in parallel utterances.

Finally, we note two other typically Austronesian features not mentioned by Klamer (2002), which are possessed by Xokleng, viz. the affixing and reduplication as productive devices, and operating basically on verbs, and the fundamental verbal distinction in both Austronesian and Xokleng between stative verbs (often translated as adjectives in English, e.g. “be frightened”, “be asleep”) and dynamic, or active, verbs.

4. Lexical Similarities

Besides Xokleng, the Kaingang language, the other member of the Kaingang language family, was included in the comparison. A 100 basic vocabulary list was compiled for Kaingang and, following the procedure described in detail in Pericliev (2006), the program was run to compare Kaingang with Malay, Tagalog, Samoan, Fijian, and Hawaiian. All pairwise comparisons, analogously to those with Xokleng, turned out to be significant, but we need not go into details here.

The procedure showed a strong relationship of the Kaingang languages with diverse Austronesian languages, but was not suited for isolating cognate words, since the computer program was run with meaning- and phonetic-similarity criteria that were much too restrictive (e.g., the use of the very same phonetic similarity criteria

in all comparisons does not take into account sound changes that as a rule take place in languages; also no meaning shifts, which are quite common in related languages, were allowed in order to preserve the statistical impartiality of the method). I therefore addressed the standard linguistic task of finding cognates. By “cognates” we mean wordforms with similar meanings from different, but related, languages, the relationship being proven by the possibility to derive the wordforms from one another, following regular sound correspondences holding between the languages. The preliminary computational approach suggested as a prospective candidate on the Austronesian side the Polynesian family (comprising the languages in the triangle New Zealand, Hawaii and Easter Island), which is furthermore the “closest” geographically. Our data for Xokleng comes from different sources (Gensch 1908; Henry 1935, 1948; Gakran 2005), and for Kaingang from the dictionary Wiesemann 2002 and Wiesemann 1972. The data on cognate sets for Polynesian are based on the *Maori-Polynesian Comparative Dictionary* by Tregear (1891), but Williams 1957, Pukui & Erbert (1986), Andrews (1865) are also consulted. The search was based on the sound changes that are known to hold in the Polynesian languages. Table 2 shows the familiar sound correspondences (e.g., Briggs 1973, Tregear 1891) in the Polynesian languages Māori (New Zealand), and Hawaiian, alongside with those we found to hold for the Kaingang family.

Table 2. Sound Correspondences between Xok(leng), Kai(ngang), Māo(ri), and Haw(aiian).

| <i>Xok/Kai/Māo/Haw</i> | <i>Examples Nos.</i> |
|------------------------|----------------------|
| p/p/p/p | 23,24,25 |
| t/t/t/k | 31,32 |
| k/k/t/k | 1,2,5,11,12,26 |
| k/k/k/? | 9,10,13,17,27 |
| k/k/h/h | 15,16 |

| | |
|-------------------|----------------------------|
| h/h/wh/h | 1,3,4,21 |
| m/m/m/m | 17,29 |
| ŋ/ŋ/ŋ/n | 11,12,24 |
| w/v/w/w | 28,30 |
| l/r/r/l | 1,2,6,7,8,9,15,16,18,19,22 |
| a/ã/ã/ã | 1,3,7,8,20 |
| a,ã/a,ã,ẽ/a/a | 5,9,11,12,24,25,26 |
| æ/a,ã/au,ao/ao | 29,30 |
| e/e/e/e | 1,2,17 |
| i/i,ĩ/i/i | 1,2,4,14,28 |
| ɔ/ũ/ũ/ũ | 18,19,23 |
| o,ɔ,u/o,a,u,õ/õ/õ | 6,10,13,27 |
| u,ɔ/u,o/u,o/u,o | 4,9,11,12,15,16,17,21,22 |

Table 3 lists 32 putative cognate sets, which are illustrations of the sound correspondences. Each cognate set has a gloss, giving a general meaning, which may be further specified for some languages at the bottom of the table if somewhat different from gloss. A dash “-” indicates a relatively clear word division in the Kaingang languages, segmenting stem from morphological endings (-*m*, -*n*, -*r*, -*y*, etc.). Brackets “()” enclose forms that are not part of the comparison (e.g. the verbal forms *ke*, *he* ‘to say’ in the Kaingang languages, which, when parts of larger verbs as in entries Nos. 18–21, Table 3, simply indicate direct speech). A slash “/” indicates doublets (e.g., the forms *kur/u* ‘cloth, blanket’ (Nos. 15, 16) are given as *kur* in Wiesemann 2002, but the final *u* is actually heard, but usually not written in Kaingang, as it reflects a regular phonological rule, adding the same or more central vowel after word final *r*, *v*, *y*); the investigator of Xokleng Henry (1935, 1948), in contrast, writes the corresponding words as *kulu*).

As seen from Table 3, the sounds (vowels and consonants) of each Kaingang-family wordform, with only a tiny number of

exceptions, are totally predictable from the Polynesian-family wordforms, according to the correspondences from Table 2.

Table 3. Putative Cognate Words Illustrating Sound Correspondences between the Kaingang and Polynesian Families

| Glosses | Kaingang family | | Polynesian family | |
|------------------------------------|--------------------|------------------------|-------------------|-------------------|
| | <i>Xokleng</i> | <i>Kaingang</i> | <i>Māori</i> | <i>Hawaiian</i> |
| 1. <i>be same or similar</i> | halike | hā ri /ke ^a | whārite | hālike |
| 2. <i>be same or similar</i> | like | ri ke | rite | like |
| 3. <i>prefix</i> ^b | ha- | hā- | whā- | hā- |
| 4. <i>whistle</i> | hui | | whio | hio |
| 5. <i>near</i> | | kakā | tata | kaka ^c |
| 6. <i>ant/insect</i> | ɔ ^d | ro ^e | rō ^f | lō ^g |
| 7. <i>sun</i> | la | rā | rā | lā |
| 8. <i>day</i> | la | kurā | rā | lā |
| 9. <i>light/glow</i> | kulaŋ ^h | kurā | kura | ʔula |
| 10. <i>stick</i> | kɔ | ka | kō | ʔō |
| 11. <i>sick</i> | kɔŋɔ | kaŋa | tuŋa/ŋoŋ o | kuna |
| 12. <i>grub</i> | kɔŋɔ | koŋa | tuŋa ⁱ | |
| 13. <i>penetrate</i> | ko | | kō | ʔō |
| 14. <i>in</i> | ki | ki | ki | ʔi |
| 15. <i>cloth</i> | kulu | kur/u | huru ^j | hulu |
| 16. <i>blanket</i> | kulu | kur/u | huru ^j | huluhulu |
| 17. <i>broth/food</i> | | kome | kome | |
| 18. <i>shake</i> | | rū-m rū-m | rūrū | lūlū |
| 19. <i>scatter</i> | | rū-m (ke) | ru ⁱ | lū |
| 20. <i>breathe with difficulty</i> | | hā-m hā-m (ke) | hāhā | hāhā |
| 21. <i>blow</i> | | hu (he) | | hu |
| 22. <i>round/roll</i> | | ror | (pi)rori | loli |
| 23. <i>birth/origin</i> | pɔ | | pū | |

| | | | | |
|----------------------------|-------|------------------|------|------------------|
| 24. <i>throw</i> | pãŋ | pẽŋ | paŋa | pana |
| 25. <i>hand/touch</i> | | pẽ | pa | pã |
| 26. <i>finish</i> | ka | ka-r | | ka |
| 27. <i>dig</i> | | kõkõ -m | kõkõ | ʔõʔõ |
| 28. <i>turn, reverse</i> | | vĩĩ-n | | wili |
| 29. <i>carry</i> | | ma | mau | mao |
| 30. <i>tree, forest</i> | wæ | vã-ŋ | wao | wao |
| 31. <i>the (def. art.)</i> | te | | te | ke |
| 32. <i>moist</i> | tuy/u | tuy ^k | tõi | kõi ^l |

^a *hã ri* and *hã ri ke* are doublets (Wiesemann 2002 and pers.com.).

^b Prefixal particles with identical form and function in present context, variously designated in the different languages: ‘emphatic’ (Xokleng), ‘assertive’(Kaingang), and ‘causal’ (Polynesian) (from *whaka/haka*).

^c ‘fruits growing in clusters’ (other Polynesian languages like Tongan, Marquesan, etc. corresponding cognates mean ‘near’).

^d Meaning ‘ant’ as Tongan *lo* and Samoan *loi*.

^e ‘small bee’.

^f ‘stick insect’.

^g ‘species of bug’.

^h Meaning ‘tomorrow’, ‘morning’ (vs. ‘evening’), ‘early’ (vs. ‘late’).

ⁱ Other Polynesian languages like Samoan, Tongan etc. also have cognate forms meaning ‘grub, maggot’ and ‘sick/sore’.

^j ‘dog skin mat’, ‘coarse hair’, etc.

^k ‘trickle intensely’.

^l ‘flow, spurt’.

The large number of shared cognates (and I have so far found well over 100 putative cognates, some of which, however, need a more in-depth analysis to see the connection), as well as the remarkable match of the sound correspondences between the Kaingang and the Polynesian languages seen in Table 2 are highly unlikely to be chance (which was illustrated by direct computations previously, cf. Pericliev 2006), and indicate strongly some

relationship between these families.

As additional level of evidence, strong even if considered in isolation, we might mention the following facts. For the Polynesian languages, it is known (Tregear 1891) that identical or related forms, see entries Nos. 11 and 12 from Table 3, are used to designate both grubs, maggots or burrowing insects and sickness, the former believed by Polynesians to be the cause for human wounds or illnesses. Perfectly analogously, in the Kaingang languages identical or related forms—cognate with those of Polynesian—are used to designate both grubs, maggots or burrowing insects and wounds or sickness (Gensch 1908, Wiesemann 2002), this society apparently sharing the same belief. A further example of similar “unusual” polysemy would be the following: Gensch 1908 lists Xokleng *lo* as meaning both ‘ant’ and ‘chin’ (Table 3 gives ɔ ‘ant’ following Henry 1935, the former author not distinguishing between *o* and ɔ; Kaingang has respectively the corresponding *ro* ‘small bee’ and *ra* ‘chin’). Significantly, Hawaiian has a similar polysemy. Thus, Andrews 1865 gives *lo* as meaning ‘a species of bug, long and with sharp claws’ and ‘the fore part of the head’, and Pukui & Erbert 1986 give *lō* ‘black insect, earwig (Dermaptera)’ and ‘front half of the skull’. Another putative example, not appearing in Table 3, would seem to be Xokleng *kɔi ka* with meanings ‘people, relatives’ and ‘sky’ (Henry 1935) (cf. also the corresponding Kaingang *kaɣkã* meaning ‘family’ and ‘sky’, Wiesemann 2002), and Hawaiian *kakai* signifying ‘a company traveling together/a family, including servants, dependents’ and ‘a cloud that hangs low near the ground’ (Andrews 1865) with similar remarkable, though indeed not exactly coinciding, polysemy.

As another piece of compelling evidence even if considered in isolation, cf. entries Nos. 1-3 from Table 3. These three entries illustrate similar composite words, e.g. Xokleng *halike*, Hawaiian *hālike*, etc., all having an identical meaning, viz. ‘be same or similar’. These composite words comprise prefixes (with no special

meaning in this context) and stems (meaning ‘be same or similar’), which are also formally similar, e.g. Xokleng Prefix: *ha* + Stem: *like*, Hawaiian Prefix: *hā* + Stem: *like*, etc. Such a coincidence is so highly unlikely, as to be practically impossible to occur by mere chance.

With the assurance of the non-fortuitousness of the evidence presented in Table 3, I give several additional putative cognates that only slightly relax the strict formal and semantic matching conditions employed above. All words below have the general meaning of the gloss unless specified otherwise; some comments in brackets may accompany the examples. The notation is the same as that used in Table 3.

Snow: Xokleng *kukrule* (Gensch 1908), Kaingang *kukr̥/ʔ* (*kɔ*), *kukr̥/ʔ* ‘ice’ = Māori *hukarere*. Xokleng’s phoneme /l/ is realized as the allophones [r] and [l], which Gensch does not distinguish, i.e. *kukrule* = *kuklule*, thus matching with the Māori word in all sounds but the word-internal *u*. The Māori word-internal *a* in *hukarere* seems to be elided in the Kaingang languages in polysyllabic words, a phenomenon observable also in the next example in Xokleng.

Smell, odour: (*di?*)*kukræ* ‘stink’ (Gensch 1908), Kaingang *kāhɔr/ɔ* ‘odourless’, *kāŋ* ‘smell’ = Māori *kakara*, Hawaiian *ʔaʔala*.

Waste water: Kaingang *ēkɔ-r* ‘sour water’ = Māori *ehu* ‘muddy’, Hawaiian *ehu* ‘dusty, disturbed’. In other Polynesian languages, these words may also refer to water and for Tahitian e.g. Tregear 1891 lists *ehu* as meaning ‘discoloured, as water by reddish earth; muddy or disturbed water’.

Salty water: Kaingang *kayã* ‘salty’ = Māori *tai*, Hawaiian *kai*. Other words for, and connected with, water also seem to have parallels in the Kaingang and Polynesian families, but this interesting topic requires more attention that cannot be paid to here.

Dirty, filthy: Xokleng *kæ-we* (Gensch 1908; segmentation in original), Kaingang *kavɜ-y* = Māori *hawa*, Hawaiian *hawa*.

Kill: Xokleng *pat̥* ‘pierce’ = Māori *patu*, Hawaiian *paku*.

Hawaiian's cognate *pa+ku* has seemingly unrelated meanings to 'kill/pierce', however cf. its component parts: *pa* 'to shoot or throw as an arrow of sugar-cane' and *ku* 'to pierce, as a spear'. In another Polynesian dialect, Paumotan, the corresponding cognate *patu* covers both meanings and signifies either 'prick' or 'kill'.

Thunder: Xokleng *tĩĩĩl* (Henry 1945), Kaingang (*ta*) *tĩ tĩrĩrĩ* (*he*) 'to thunder' = Māori (*wha*)*titiri*, Hawaiian (*he*)*kili*. Other Polynesian dialects have similar forms, cf. e.g., Samoan (*fāi*)*titili*, Tuamoto (*fa*)*tutiri*, Tahiti (1773) (*pa*)*tiree* 'it thunders', etc. Henry suggests that the Xokleng word is onomatopoeic; whether this is indeed the case or not, however, is immaterial in the present context insofar as the Kaingang and Polynesian languages do have similar forms, while other, non-Kaingang or Polynesian languages, would have quite different forms.

5. Conclusion

In this paper, I have shown linguistic evidence, especially of lexical nature, that seems to corroborate the idea of affinity between the Kaingang language family, in South-eastern Brazil, and the Polynesian language family quite strongly. The number of reasonably convincing cognates found, as well as the remarkable sound matches between Kaingang and Polynesian attest to this conclusion. Taking into account all linguistic evidence accumulated so far, this relationship may probably be plausibly explained by common descent from the same (higher-order) language family, the exact nature of which still remains to be established (as a very tentative guess, not improbably Proto-Eastern-Oceanic, including, besides Polynesian, Fijian, Rotuman and certain languages of the Solomon-New Hebrides chain). Some recent population genetic investigations have found the Macro-Ge-speaking Xikrin and the Tupi-speaking Parakanã (Tupi is believed to be related to Macro-

Ge) to be genetically similar to Indonesians and South-East Asian populations (cf. Ribeiro et. al. 2003 and references therein), which also leads in a similar direction. Our result raises a number of further problems, in both the study of Oceanic and Macro-Ge languages, and specifically Kaingang, as well as such to be addressed by specialists in other fields like anthropology, genetics, and archeology. I hope that the present study may stimulate some further efforts in this direction.

References

- Andrews, L. 1865. *A Dictionary of the Hawaiian Language, to which is Appended and English-Hawaiian Vocabulary and a Chronological Table of Remarkable Events*. Honolulu, HI: Henry M. Whitney.
- Briggs, B. 1973. The Languages of Polynesia. *Current Trends in Linguistics* 466-505. The Hague: Mouton.
- Gensch, H. 1908. Wörterverzeichnis der Bugres von Santa Catharina. *Zeitschrift für Ethnologie* 40, 744-759.
- Gordon, G. (ed.). 2005. *Ethnologue: Languages of the World, Fifteenth Edition*. Dallas, TX: SIL International.
- Henry, J. 1935. A Kaingang Text. *International Journal of American Linguistics* 8, 172-218.
- _____. 1948. The Kaingang Language. *International Journal of American Linguistics* 14, 194-204.
- Klamer, M. 2002. Typical Features of Austronesian Languages in Central and Eastern Indonesia. *Oceanic Linguistics* 41, 363-383.
- Murdock, P. 1970. Kin Term Patterns and Their Distribution. *Ethnology* 9, 165-207.
- Pericliev, V. 2006. Significant Lexical Similarities between a Language of Brazil and Some Languages of Southeast Asia and Oceania: From Typological Perspective. *Journal of Universal Language* 7.1, 121-145.
- Pukui, M. & S. Erbert. 1986. *Hawaiian-English, English-Hawaiian Dictionary*. Hononulu, HI: University of Hawaii Press.
- Ribeiro, M., S. Figueiredo & F. Sonati, 2003. Haplotypes of α -Globin Gene Regulatory Element in Two Brazilian Native Populations. *American Journal of Physical Anthropology* 121, 58 - 62.

- Ruhlen, M. 1987. Classification. *A Guide to the World's Languages* 1. Stanford, CA: Stanford University Press.
- Tregear, R. 1891. *Maori-Polynesian Comparative Dictionary*. Wellington: Lyon & Blair.
- Urban, G. 1985. Ergativity and Accusativity in Shokleng. *International Journal of American Linguistics* 51, 164-87.
- _____. 1986. Semiotic Functions of Macro-parallelism in the Shokleng Origin Myth. In J. Sherzer & G. Urban (eds.), *Native South American Discourse* 15-57. Berlin: Mouton de Gruyter.
- Valdés-Pérez, E. & V. Pericliev. 1999. Computer Enumeration of Significant Implicational Universals of Kinship Terminology. *The Journal of Comparative Social Science* 33, 162-174.
- Wiesemann, U. 1972. Die Phonologische und Grammatische Struktur der Kaingáng-sprache. *Janua Linguarum Series Practica* 90. The Hague: Mouton.
- _____. 2002. *Dicionário Bilingüe Kaingang - Português*. Curitiba: Editora Evangélica Esperança.
- Williams, W. 1957. *A Dictionary of the Maori Language*. Wellington: Government Printer.