Sebastian Drude[1]

On the Position of the Awetí Language in the Tupí Family

1. The Awetí and their language

The Awetí are an indigenous group living in the Xingú Park in Mato Grosso, in central Brazil. They are part of the cultural area of the Xingú Headwaters ("Upper Xingú"), well known for its remarkable combination of cultural homogeneity with considerable linguistic diversity: there are at least eight major dialects of six different languages belonging to four different linguistic genetic groupings.

Various factors including the oral tradition of the Upper Xinguan peoples indicate that the ancestors of the current Awetí arrived in the region around 1700.[2] They possibly were part of several Tupí speaking groups that joint a cultural system founded by Arawak-speaking groups (around 900 DC) which already managed to integrate Karib-speaking groups from 1400 DC on. At the same time as the Awetí, or somewhat later, other Tupian groups arrived and later merged into what today are the Kamayurá. At last the Trumai (speaking an isolate language) arrived, shortly before the first pacific contact with non-Indians at the end of the 19th century.

The Awetí of today are the result of the pre-historical fusion of at least two groups: the Awetí proper (***Awytyza 'ytoto***, in their own language), and the Enumaniá, both allies of the (Karib-speaking) Bakairí which today are located outside the Upper Xingú, to the south-west. The Awytyza were integrated into the Upper Xinguan society first, but remained allied with the Enumaniá, which were 'wild' from the point of view of the Upper Xinguan. When the men of the first were almost all killed by the Tonoly (a non-Xinguan tribe, possibly a subgroup of the Kayabí), the Enumaniá took their place in the Upper Xinguan system. So the Awetí hold that they are indeed the descendants of the Enumaniá rather than of the Awytyza, and their language is that of the Enumaniá.[3]

---

[1] Freie Universität Berlin and Museu Paraense Emílio Goeldi. The work underlying this paper has been carried out within the Tupí Comparative Project headed by Denny Moore, especially in cooperation with Sérgio Meira.

[2] Cf. various contributions in FRANCHETTO and HECKENBERGER (2000).

[3] For a detailed account of the historical tradition of the Awetí, see SOUZA (2000).

The German explorer and anthropologist KARL VON DEN STEINEN was the first to visit the region, in 1884 and in 1887. At the occasion of his second trip, he visited the Awetí and stated that they were well introduced into the Upper Xinguan society (STEINEN, 1894). There was at least one village, possibly of the average size of a Xinguan village to that time, which apparently was around 400–500 souls. (Archaeological studies show that before the 16th century there were villages of several thousands inhabitants, cf. HECKENBERGER (2000)).

In the first half of the 20th century the Awetí, like the other Upper Xinguan groups, were (again) dramatically reduced in number, principally by diseases. There is an official census dating from 1954 that gives a number of 23 Awetí only. Today there are more than 140 Awetí living in two villages, after a split of the group around 2002. The main village is almost at the same place were the Awetí were visited by VON DEN STEINEN 120 years ago. The second, newer and smaller one is about 20 km north, also on the right margin of the Tuatuari stream, at only some 5 km distance of the 'Posto Leonardo'. In this second village Awetí may be suppressed by Kamayurá as the first language of the majority, but in the main village still almost all children continue to acquire Awetí as their first language.

The Awetí and their language arguably continue to be the least studied in the Upper Xingú, despite research done by RUTH MONSERRAT in the sixties and early seventies of the 20th century. The research by the author of the present contribution started in 1998. From 2000 on it is included in the research programme DOkumentation BEdrohter Sprachen (Documentation of Endangered Languages, DOBES, funded by the German Volkswagenstiftung), focussing on documenting the language and aspects of the culture.

2. The question of the genetic position of Awetí

The first who registered the Awetí language, Karl von den Steinen (1894), already stated that this language is a Tupian language. To that time (end of séc. XIX), no internal subdivisions of this large language family were established. Languages of the widespread Tupí-Guaraní branch, however, were well known due to the early contact of the European with the Tupinambá living along the Brazilian cost, and with the Guaraní in the Paraná-Paraguay basin. On the basis of their languages also emerged 'linguas gerais' that were used by the non-indigenous population of southern Brazil until the 18th century and in Paraguay and in parts of the Amazon region even until today. So, for von den Steinen and

his contemporaries, apparent similarities of Awetí with these languages were sufficient to establish that Awetí belongs to this language family.

In the second half of the 20th century, especially when the Tupian languages of Rondônia got known, it was discovered that the Tupian languages indeed represent a large and quite heterogeneous language family (a language 'stock', in some terminologies) of formerly about 45 languages, where the mostly better-known languages closely related to old Tupí and old Guaraní (ca. 20 languages spoken by more than 40 ethnic groups) are just one branch, or sub-family, Tupí-Guaraní (TG) among others. RODRIGUES (1964) established seven sub-families of Tupí. Based on nothing else then the few word lists of the early explorers, he put Awetí inside the Tupí-Guaranian branch, close to Kamayurá (in this, following others before him), probably due to the close geographic localization of both languages.

The first linguistic research on Awetí was carried out by the Brazilian linguist RUTH MONSERRAT from 1967 to 1975. Her findings revealed too many differences from the typical Tupí-Guaranian languages for considering Awetí a member of this sub-family. When RODRIGUES published his internal classification of the Tupí-Guaranian family (1984/85), he classified Awetí as a one-member sub-family inside the large Tupí family, which now is seen to consist of ten branches: Tupí-Guaraní, Awetí, Sateré-Mawé, Mundurukú, Jurúna, Tuparí, Arikém, Mondé, Ramaráma, and Puroborá.[4] At the same time, it was recognized that Awetí, similar to Sateré-Mawé (also a branch consisting of one single language, and also formerly considered to be part of Tupí-Guaraní), had much more similarities with the Tupí-Guaranian subfamily than any language of the other branches.

---

[4] Only about 14 non-Tupí-Guaranian Tupían languages are currently used on a daily basis. All are endangered to disappear within the next 100 years. In the Tupí Comparative Project, at least one language of each branch of the Tupí family is studied by a member of the project: The little that can be recovered of the Puroborá language by ANA VILACY GALÚCIO; Karo, arguably the only Ramarama language, by NILSON GABAS JÚNIOR; of the Mondé languages Suruí by MARIANA LACERDA and Gavião (and Mondé in general) by DENNY MOORE; the only surviving language of the Arikém family, Karitiána, by LUCIANA STORTO; of the Tuparí family: Mekéns by ANA VILACY GALÚCIO and Ayuru by DIDIER DEMOLIN; of the Juruna family, Xipaya is studied by CARMEN RODRIGUES; the two languages of the Mundurukú family, Mundurukú and Kuruaya, by GESSIANE PICANÇO; the Sateré-Mawé language by SÉRGIO MEIRA and the Awetí language as well as Paraguayan Guaraní (Tupí-Guaraní) by the author, SEBASTIAN DRUDE. A first result of the common work is the probable subgrouping of Karo (Ramarama) together with the almost extinct Puroborá, see GALUCIO and GABAS JR. (2002).

In other words, there is the hypothesis of a more comprehensive branch of the Tupí linguistic family which includes three sub-branches: The large Tupí-Guaraní family, and each of the two single languages Sateré-Mawé and Awetí. At some point, SÉRGIO MEIRA and the author began to call this intermediate Tupian sub-family 'Mawetí-Guaraní', using elements of the names of all three branches. We will continue to use this name in this paper.

Our findings in the Tupí Comparative Project (TCP) strongly corrobate the hypotheses of the existence of this sub-family 'Mawetí-Guaraní'. To begin with, we found much more shared cognates among the three branches of Mawetí-Guaraní (using the same lexical lists for all languages, as far as possible). This tendency is confirmed by the Swadesh list of 100 diagnostic words. The result of a study of the TCP (MEIRA et al., in prep.) is that the number of cognates between arbitrary Tupian languages belonging to different families is around 30, while the number of cognates within the languages of Mawetí-Guaraní are around 50 and higher, and within the Tupí-Guaraní family around 70 and higher. Though we members of the TCP generally are sceptical about the premises of lexicostatistics and, in particular, glottochronology, we assume here that these are certainly significant numbers.

To discuss the reasons for assuming Mawetí-Guaraní in detail is beyond the scope of this paper. The question that is in the focus here is the internal grouping, if there is any, within this sub-family. There are four logically possible configurations. These configurations can be represented by tree-diagrams as in Table 1.

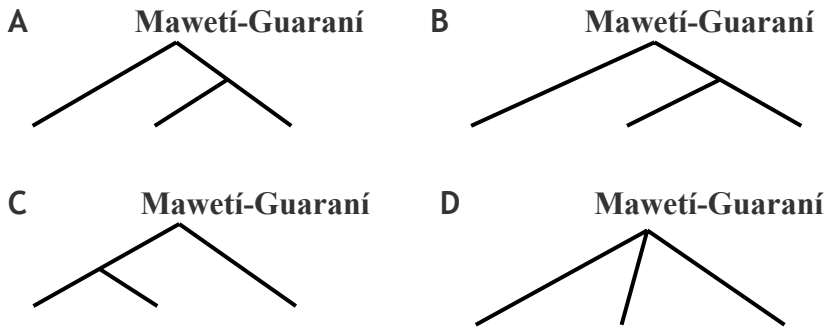
The most common superficial impression seems to be that Awetí is somewhat closer to the Tupí-Guaraní family than Sateré-Mawé is (configuration A). Again, the study applying lexicostatistics points in the same direction: Awetí shares a middle of 54,5 cognate words in the Swadesh-list with the four TG languages studied while Sateré-Mawé has a middle of 48,75 with the four TG languages. But these close numbers should not be taken more than a weak indication.

But perhaps this impression is misleading and indeed the ancestors of the Awetí separated from the ancestors of the Sateré-Mawé and Tupí-Guaraní before these latter separated one from another (configuration B). Still, it may be even wrong to follow the traditional, somewhat Tupí-Guaraní-centric approach to ask 'what is closer to Tupí-Guaraní' – maybe rather Awetí and Sateré-Mawé constitute a branch of Mawetí-Guaraní, both being genetically closer one to another than to the Tupí-Guaranian languages (configuration C). Note that Awetí

has with Sateré-Mawé 51 cognates in the Swadesh-list, which is, oddly enough, higher than would be expected if Sateré-Mawé split off first. The differences are, however, certainly too low to have any deeper significance.

So, finally, maybe it is just impossible to establish any subgrouping inside the Mawetí-Guaraní family, all three branches being more or less equidistant one of another (configuration D). This would allow an interpretation in historical terms yielding the hypothesis that the separation of the ancestors of the speakers of the three branches took place simultaneously or during a quite short period of time (at least, too short for any significant linguistic changes to have occurred between the two splits).

Table 1 : Possible genetic relations within Mawetí-Guaraní



3. DIETRICH's (1990) study

The only study based on language data that deals with the internal grouping of the Mawetí-Guaraní family until today is that of DIETRICH (1990).[5] In fact, its focus is on the internal grouping of a representative sample of languages and dialects of the Tupí-Guaraní family, but the two other languages of Mawetí-Guaraní, Sateré-Mawé and Awetí, have also been included as far as possible.

DIETRICH's study, however, is not of the historical comparative type *strictu sensu*; he does not aim at a reconstruction of the hypothetical proto-language (proto-Tupí-Guaraní, pTG, or even proto-Mawetí-Guaraní, pMATG)

[5] The earlier work of LEMLE (1971) did not consider Awetí (nor Sateré-Mawé), probably due to lack of data.

and at the identification of the historical sequence of changes the individual languages undertook each after its respective separation. He rather established six sets of diagnostic parameters (three phonetic-phonological and three morpho-syntactic ones), examined the value for each language with respect to the given parameters ('+' '–' '±', or '?') and then applied a statistical calculation that represents the overall similarity of the languages among each other (a sum of shared values of features). As a result, each pair of languages has a number that represents their 'proximity' (statistical similarity) according to the chosen parameters. He presents the results in tables and also in form of two-dimensional graphics where the languages are posited nearer or farer according to their statistical similarity (number of shared characteristics with respect to the chosen parameters).

For a valuable appreciation of DIETRICH's work, see JENSEN (1998). We support here suggestions for refinements but still hold that a study of this type can only give supporting evidence and additional interesting information to a rigorous study applying the historical comparative method.

In the case of Awetí, DIETRICH did not know the correct value of several of his features, so his results for this language have to be taken *cum grano salis*. Here, we first repeat the results of DIETRICH's study and then show our own findings with respect to his parameters.

DIETRICH's first graphic represents the 'proximity' of the languages of his sample according to his phonetic-phonological parameters, repeated below as Table 2.[6] He identified two main subgroups. The first one includes the 'average' Tupí-Guaranian languages while the second one is restricted to some deviant languages (DIETRICH uses the term 'innovative', which may be true for several of his parameters.) Awetí belongs to the second division, together with Guayakí, Xetá and Cocama. These latter are precisely three Tupí-Guaranian languages which are suspect to show strong substrate influence, that is, to be languages that

[6] Lines between language name abbreviations indicate close proximity between the languages. The abbreviations for language names used are: Ap: Apapocuva, AsT: Asuriní of Trocará, AsX: Asuriní of Xingú, Av: Avañe'e (Guaraní), Aw: Awetí, Ch: Chiriguano, Cn: Canoeiro, Co: Cocama, CT: Chiriguano-Tapyj, Em: Emérillon, Gj: Guajajara, Gk: Guayakí, Gy: Guarayo, Jm: Juma, Kb: Kayabí, Km: Kamayurá, Kw: Kaiwá, Mb: Mbyá, Ps: Pauserna, Pt: Parintintin, Si: Siriono, SM: Sateré-Mawé, Tb: Tupinambá, Te: Tembé, Tp: Tapirape, Ub: Urubu-Kaapor, Wa: Wayãpi, WJ: Wayãpi/Jari, Xt: Xetá. The new position of Awetí according to our results is included in the Tables and is marked in italics and underlined.

are spoken by groups that originally spoke other languages and then eventually took over a Tupí-Guaranian language.

Table 2 : DIETRICH's (1990: 59) phonological grouping of Mawetí-Guaraní

Southern Group			*Amazonian Group*		
Ap – Mb				**Em – Cn**	
		Sn			
Gy			**Wa**	**AsT – AsX**	
			Wj Kb		
Kw		**Pt – Jm**			
| – Tb			**Gj –**	**Km – Tp**	
Av	**CT**		**|**		
	|		**Te**		
I	**Ch**	**Ub**			
		Ps			
		Aw	**SM**		
				Aw	
II					
Gk	**Xt**			**Co**	

Within division II, Awetí is somewhat closer to division I than are the other languages. Awetí is also located to the very right end. This is to indicate that it shows much more affinity to DIETRICH's 'Amazonian Group', i.e. to a group of languages that show more similarity among each other than to the languages of the `Southern Group' (basically the languages or dialects of the Guaranian cluster, together with Tupinambá). It is remarkable that, according to DIETRICH's results, Sateré-Mawé is still included within division I (even if marginally so) while Awetí is not.

In fact, DIETRICH did not know the following properties of Awetí with respect to his diagnostic phonological parameters: 1) Awetí verb forms and many noun forms usually occur with unstressed suffixes (as he assumes for Parintintins), so a '±' seems more adequate than his '+' for 'stress on last syllable'.[7] 2) As far as we can see, Awetí has almost no syncope, so we give '–' for 'strong syncope', '±' for 'slight syncope' and '+' for 'no syncope'. 3) There

[7] This, however, may hold for several Tupí-Guaranian languages as well if occurrences in texts are analysed.

is the nominal past/perfective suffix ***-put***, arguably a reflex of proto-Mawetí-Guaraní **-**, so his feature 'reflex of PTG ***-p***is ***- (r)***' should get a '+' and the feature '…is ***-k***' will get a '–'. 4) At least for the examples he gives, Awetí has – phonetically – a reflex of ***/t/+/i/***, so we give a '+' for his feature 'či', besides a '+' for his feature 'ti', which correctly reflects that phonologically Awetí has indeed preserved ***/t+i/***.

As a result, Awetí has more in common with the languages of the Guaranian group and especially with Guarayo than it had according to DIETRICH's calculation, but somewhat less with the northern languages and, in particular, with Siriono.[8] Nevertheless, Awetí continues to be statistically closer to the Amazonian languages. But his classification into the group of 'most innovatory' languages does not seem justified any more, as its overall score is considerably higher now, even higher than that of Sateré-Mawé, which DIETRICH includes in the 'less conservative' languages. We therefore put Awetí (new position in italics and underlined) more to the left and above the dividing line in Table 2.

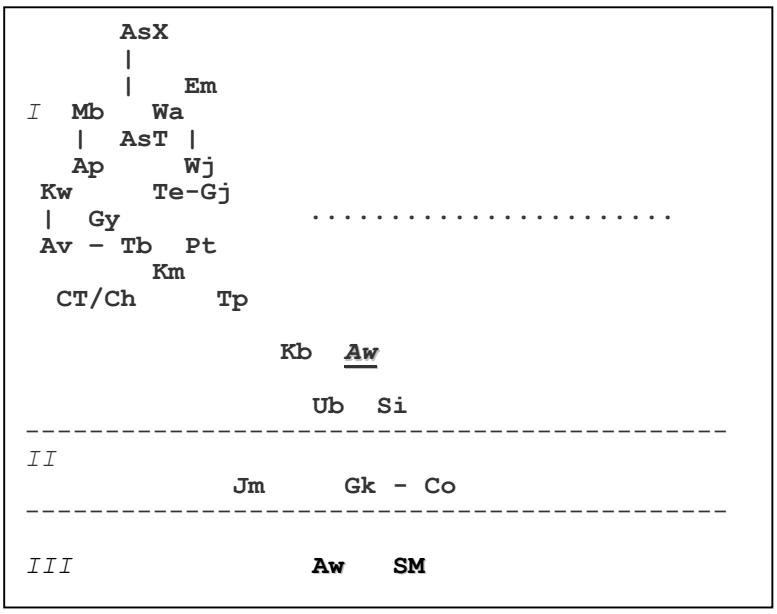
In a second step, DIETRICH used a set of 'grammatical and morphological' criteria to establish a similar grouping of the languages of his sample. We repeat his summarizing graphical representation below in Table 3. He found three main divisions among the languages according his grammatical criteria, where the large majority belongs to the division I, Juma, Guayakí Cocama belong to division II and only Awetí and Sateré-Mawé constitute division III. So with respect to division III, DIETRICH's results correspond to the genetic classifications proposed by others (see last section, above) which posits these two languages in separate branches, apart of Tupí-Guaraní. His focus, however, is on conservativeness compared to Tupí-Guaraní, so we should not interpret his findings as indicating Awetí and Sateré-Mawé being very close one to another, but just that both are statistically distant from the average Tupí-Guaranian languages.

The distinction between 'Amazonian' and the 'Southern' languages has not been so clear to DIETRICH using his grammatical criteria.

---

[8] The new proximity values for Aw are: 26: Kb,Gy; 25: Ap,Cn,AsX; 24,5: Tb,SM; 23,5: Pt; 23: Te; 22,5: Wa; 22: Mb, Gk,Co; 21,5: Av,Ub,Jm; 21: Kw,Em; 20,5: Gj,Tp,Km; 19,5: WJ; 19: AsT; 18: CT; 17,5:Ch,Si; 13: Xt. Summing these values gives an overall score of 636,5, in comparison to 617 according to DIETRICH's results and 625,5 for SM.

Table 3 : DIETRICH's (1990: 98) grammatical grouping of Mawetí-Guaraní



With respect to the grammatical parameters, DIETRICH's data on Awetí were much more deficient than in the phonological case. If we complete his data, the results for Awetí change considerably.

DIETRICH's first set of grammatical/morphological criteria are complete and correct from our point of view. In his second set, the following information was missing: 1) the predicative negation in Awetí, ***-ka***, does not seem to be cognate with pTG ***-(r)i*** ('–' for 'Neg'). 2) There is no negating suffix ***-ã*** in Awetí ('–' for '-ã'). 3) There is no cognate of a ***ne*** suffix or particle for future in Awetí ('–' for 'ne'). 4) Nominal tense exists ('+' for 'NT', although these categories have more than just temporal semantic effects). 5) There seems to be no cognate of a ***pir*** nominalising suffix in Awetí ('–' for 'pyr'). 5) There is a cognate of the ***emi-*** nominalising prefix in Awetí ('+' for 'emi'). 6) The cognates of the prefixes ***je-*** and ***jo-*** exist in Awetí (***te-*** and ***to-***, '+' for 'je' and 'jo').

Most of the features unknown to DIETRICH's in his third grammatical set do positively hold (are marked with a '+') for Awetí: 1) the suffix ***-uka*** exists in Awetí ('factitive voice'). 2) The prefix ***-ẓo***, cognate to ***-ro***, exists in Awetí ('factitive sociative voice'). 3) The prefix ***-mo*** exists in Awetí. 4) The prefix

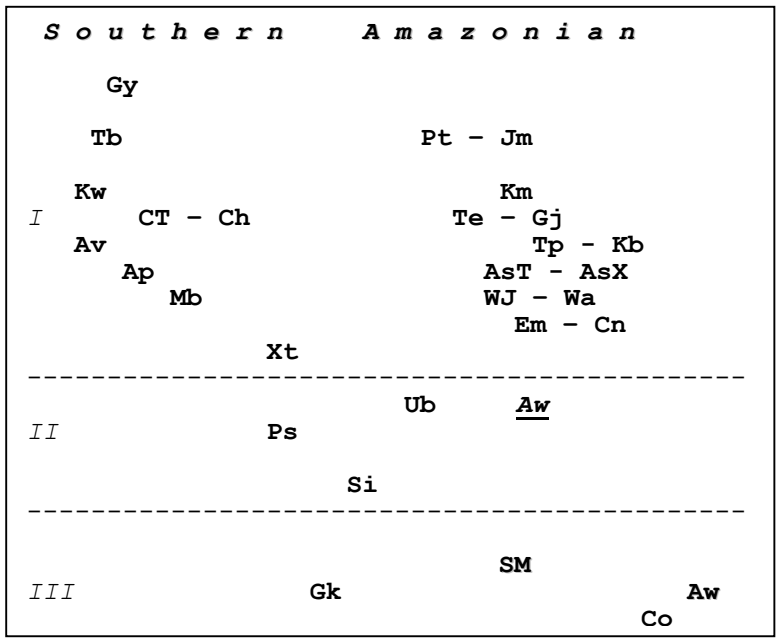
-po(r) exists in Awetí, but it is not quite a 'generic human object', so we give a '±' for this criterion. 5) Awetí does not show any special 3rd person form related to ***hese*** ('–' for '***s-ese → hese***'). 6) Awetí has the 3rd person pronoun ***nãˀ***, which could be related to DIETRICH's ***ŋa***, but only in the male variety (see next section) and only in the singular, so we give a '±' for this criterion.

According to DIETRICH's methodology, Awetí results to be much closer to the typical Tupí-Guaranian languages than he calculated.[9] Its overall score of grammatical/morphological proximity is more than 60% higher when our values are accepted. It is then similar to that of Kayabí and much higher than that of SM, so it should belong into DIETRICH's group I instead of group III. We show again a suggestion for a corrected position of Awetí in the graphic (Table 3, Awetí's new position is again marked by italics and underline).

Finally, DIETRICH combined the findings based on his two sets of criteria into a single merged statistical similarity grouping (not: a genetic classification). In this combined result, he again distinguishes three divisions. According to him, Awetí belongs again into the third, most divergent division, together with Sateré-Mawé, as was to be expected, and with Guayakí and Cocama.

[9] Our results for the proximity values are: 26,5: Kb; 25,9: Co; 25,8: Si; 24: Gk; 23,7: Pt,Ub; 23,5: AsT,Te,Wa,Wj; 23: Sm; 22,7: Tp; 22,5: Mb; 22,2: Km; 22: ASX,Gy; 21,5: Tb; 21: Ap,Kw; 20,5: Av,Gj; 19,5: Ch,Ct; 18,3: Em; 14,7: Jm. The sum ('overall score') of these figures is 580, in comparison with DIETRICH's 353,4. SM: 430,6; Kb: 590,9.

Table 4 : DIETRICH's (1990: 111) combined grouping of Mawetí-Guaraní



As can be seen by Dietrich's division III, his results have no direct implication for any truly genetic classification of Mawetí-Guaraní. Both, Sateré-Mawé and Awetí are in this most deviant division, but so are Guayakí and Cocama, certainly two 'deviant' languages, but presumably not due to genetic distance to the prototypical (central) Tupí-Guaranian languages, but to substrate influence. Genetically, they are usually still considered Tupí-Guaranian.[10]

Also, in DIETRICH's table, Sateré-Mawé and Awetí are quite close one to another, compared to Guayaqui, for instance, but this cannot be interpreted in genetic terms, nor can the close proximity DIETRICH stated between Cocama and Awetí.

If we consider our figures, the scenario changes a lot for Awetí. It is then posited more to the left (somewhat closer to the Guaranian languages) and definitively closer to the 'conservative' languages in DIETRICH's division I. His

[10] See JENSEN (1998), MELLO (2002), RODRIGUES (1984/85; 1985) and RODRIGUES and DIETRICH (1997). SCHLEICHER (1998), however, puts Cocama (and Siriono) outside TG.

overall score is now 1216,5, compared to DIETRICH's 969,5, quite similar to the result for Urubú, for instance, and should therefore be posited at the same height as this latter language, that is, in division II, not division III.

As said above, a study of this type should not be taken to give decisive answers to questions of genealogical classification (this requires, before and above all, historical comparative work), even if refinements of the method are applied and the data completed or corrected. When the study has been redone with more complete data on Sateré-Mawe, the result may give some secondary evidence for the internal sub-grouping of Mawetí-Guaraní.

When we examined DIETRICH's criterion of a 3rd person pronoun ***ŋa***, (point 6 for the third set of grammatical criteria, above), we mentioned the 'male variety' of Awetí. In the next section, we present the two Awetí genderlects examine their significance of for the question of genetic grouping.

4: The two 'genderlects' of Awetí

It seems that there are only very few minor dialectal differences within the Awetí language. It is not clear if the Enumaniá and the former Awytyza spoke already the same language, but at any rate there seem to be no or very few traces of the idiom of the latter.

There are, however, two major varieties of Awetí which can be named 'genderlects': One is used by men, the other one by women. The differences between the two varieties are not on the phonetic / phonological nor on the lexical level, but in morphosyntax, especially in the paradigms of personal and deictic pronouns and (related) prefixes of nouns.[11] The paradigms of personal pronouns are given in Table 5.

[11] For a more complete account of the genderlects, see DRUDE (2002).

Table 5 : Personal pronouns in the male and female variety in Awetí

Category	Male variety	Female variety
1. sg.	***atít — [ʔaˈʧĩtˀ]***	***itó — [ʔĩˈtɔ]***
2. sg.	***en — [ˈʔɛn]***	
3. sg.	***nã — [ˈnã]***	***ĩ — [ˈĩ]***
1.+2. pl.	***kajã — [kãˈɲã]***	
1.+3. pl.	***ozóza — [ʔɔˈzo̯zo̯a]***	
2. pl.	***e'ipé — [ʔɛʔĩˈpɛ]***	
3. pl.	***tsã — [tsã]***	***ta'í — [taˈʔi]***

The nominal prefixes that indicate person (usually of the 'possessor') differ also according to male or female variety, but only in the third person. In the case of the male variety, the prefixes are clearly related to the independent pronoun given above: ***nã-*** before consonants and ***n-*** before vowels. In the female case, if a vowel follows, the prefix ***t-*** is apparently not related, while the form before consonants, ***ĩ-***, is identical to the independent form, but sometimes the non-nasal ***i-*** is also possible.

Finally, the forms of some deictic pronouns differ as well, cf. Table 6. [12]

Table 6 : Basic deictic pronouns in the male and female variety in Awetí

Position	Male variety	Female variety
close to speaker	***jatã — [jaˈtã]***	***ujá — [uˈjá]***
close to hearer	***kitã — [kĩˈtã]***	***akýj — [aˈkɨj]***
distant of both, sp.+hr.	***kujtã — [kujˈtã]***	***akój — [aˈkɔj]***

As can be seen easily, it seems that the male forms appear to be derived from or at least related to the second syllable of the female forms, showing an additional final element ***-tã***. This element is still conspicuous as it does not cause nasalization of the first syllable as would be expected by the rules of nasal harmony in Awetí. This indicates that, in the male variety, the vowels of the first

[12] Differently from what we stated formerly (DRUDE, 2002), new data confirms that there is indeed a functional distinction between ***akýj*** and ***akýj*** in the female variety.

syllable still maintain their intrinsically oral character (as in the female variety). The element ***-tã*** shows, however, similarities to the third person pronouns ***nã*** and ***tsã*** (in the male variety), especially in having a variant with an additional final ***-n*** (***…tãn***, equally ***nãn*** and ***tsãn***) before vowels in certain contexts (especially before verb stems (with 'absolutive' function), before some postpositions, and before the very frequent clause-final particle ***a'yn***).

Due to these characteristics of the two varieties, we hesitate to attribute the genderlects of Awetí to two different substrates (e.g., Awytyza and Enumaniá). Even if this were the case, it could not be affirmed that one of the two substrate languages would be a non-Tupian, or a Tupí-Guaranian, language. They would then rather have been two dialects of the same language.

As to the internal grouping of Mawetí-Guaraní, the two varieties do not show significant differences. The deictic forms are related between the two varieties, and there are no reconstructions for genuine third person pronouns for pTG or even pMATG, so that the development of these is a shared development of both varieties. A very weak evidence is the supposed similarity of Awetí ***nã***, with DIETRICH's pTG ***ŋa***, see the last section above. On the other hand, the first person singular pronoun of the female variety seems to be cognate with the Sateré-Mawé form ***ujto*** (and possibly also with the pTG form ****itʃe***[13]), while the male form is more difficult to relate to the form of any of the languages, although its segments are all familiar (***t*** and ***i***; the ***a*** may be related to the verbal first person prefix). All these evidences considered, we conclude that there is no consistent picture as to the proximity of any of the two genderlects to the other branches.

5. First results of phonological (lexical) comparison 1: (near) identity sets

a. General remarks

As said earlier, in the Tupí Comparative Project (TCP) we are applying the historical-comparative method to languages of all major branches of the Tupian language family.[14] Especially SÉRGIO MEIRA and the author are working on a reconstruction of 'proto-Mawetí-Guaraní' (pMATG), the hypothetical protolanguage from which the languages of the Mawetí-Guaraní subfamily

[13] Reconstructed forms or sounds (here, for pTG) are marked by an asterisk ***** (which does not indicate ungrammaticality).

[14] See footnote 4.

developed. In this section we will examine first results of this joint work carried out in cooperation with the group of the TCP.

The comparison has been done between Sateré-Mawé (henceforth also SM, for short; data from field-work carried out by SÉRGIO MEIRA, in part based on earlier published sources), Awetí (henceforth also A; data from field-work by the author) and proposals for reconstructions of the hypothetical protolanguage of the Tupí-Guaraní (TG) family, 'proto-Tupí-Guaraní' (pTG), which itself developed from pMATG after the separation of the ancestor languages of Awetí and Sateré-Mawé, but before its dialects later split up in the today about 20 Tupí-Guaranian languages.[15] (If needed, the results may be refined by also considering data from individual Tupí-Guaranian languages, but we start (as usual) with comparing the hypothetical reconstructed intermediate protolanguage pTG.)

According to the comparative method, if there are words, one each in related languages, which are formally similar and have the same or a closely related meaning, it is probably the case that they originated from a single word in the hypothetical proto-language – except for the case that the word was introduced into one ore several of the languages by borrowing. If borrowing can be excluded, the set of words are considered to be *cognates*. In the case of Mawetí-Guaraní, we can base our study on more than 150 triplets of probable cognates and much more pairs of cognates between two of the three languages.

Within the cognates, one compares correspondences between the individual sounds at analogous positions and so establishes regular *sound correspondences*. Observing the environment and applying general rules of regular sound changes occurring in languages, one proceeds by proposing the correspondent sounds of the protolanguage and so reconstruct the hypothetical shape the words in the protolanguage might have had. It is important to be able to show not only resemblances but a regular way of changes for each sound and word of the daughter languages that are *reflexes* of the reconstructed sound and word of the proto-language.

The historical-comparative method can then also be applied to grammatical features (especially, reconstructing functional morphemes and some syntactical properties).

We are currently preparing a more complete publication of the cognate sets and our proposals for the reconstruction of pMATG (MEIRA and DRUDE, in

[15] Data has been taken from the proposed reconstructions presented in the literature cited in footnote 10.

prep.). Here we will restrict ourselves to results of the phonological comparison as far as relevant for the question of the genetic sub-grouping, discussing sets of corresponding sounds (with at least one illustrative example for each), but not always giving the reconstructed sounds of the protolanguage.

There are many 'stable' correspondences (where all three languages show the same sound in analogous positions). What is of interest here are cases where one language diverges from the other two or where all three have different reflexes in a way that suggests that two of the languages **share** a sound change. The more unexpected the shared change of two languages, the more probable is it that this change occurred only once, and that the two languages therefore separated one from another **after** this change occurred – that is, that they together constitute a branch, or sub-family, of Mawetí-Guaraní.

The non-identity correspondences will be discussed in the following sections. In this section, we show the stable correspondences. We start by examining the vowel correspondences.

b. Stable correspondences: Vowels

The vowels are in most cases the same in analogous positions in cognate words of Awetí, Sateré-Mawé and the reconstructed proto-Tupí-Guaraní (pTG). Where they are not, there is usually a regular explanation. We will illustrate these statements below, starting directly with the correspondences in Table 7. Almost all cognate sets given later (in different contexts) illustrate these correspondences sets. (A correspondence set, or CS, is a group of correspondent sounds, one of each language. When cited, the sounds are listed, with a colon between them, in the order pTG, A, SM.)

Table 7 : Most common correspondences between vowels

pTG	Awetí	Sateré-Mawé
***a**	***a***	***a***
***ã**	***ã***	***ã***
***e**	***e***	***e***
***ẽ**	***ẽ***	***ẽ***
***i**	***i***	***i***
***o**	***o***	***o***
***u**	***u***	***u***
***ɨ**	***ɨ***	***ɨ***
***i**	***ɨ***	***i***

Table 7 contains only two identity correspondence sets for nasal vowels, /ã/ and /ẽ/. Other (phonologically) nasal vowels occur less frequently and have therefore here been omitted. Correspondences between nasal and oral vowels with the same quality also exist, but are mostly easily explained if one takes into account the rules of nasal harmony (or nasal spreading) which exists in the three branches and probably similarly existed in the proto-language.

There is only one regular correspondence of different vowels, ***i : ɨ : i***, cf. the last line in Table 7. Here, A shows ***/ɨ/*** where the other branches show ***/i/***. This is exemplified in Table 8.

Table 8 : Examples for *i : ɨ : i*

Gloss	**pTG**	**Awetí**	**Sateré-Mawé**
chest	***potiʔa**	***poz̧ɨʔa***	***potiʔa***
mosquito	***jatiʔũ**	***taz̧ɨʔũ***	***watiʔũ***

As can be seen in Table 8, the CS ***i : ɨ : i*** involves the environment ***t__*** in pTG and SM, but in A, the environment is ***z̧__***.[16] The sound ***/z̧/*** (a retroflex voiced fricative) is unique to A among the three languages. It corresponds often to ***/r/***

[16] That is, the ***/ɨ/*** in SM and pTG occurs after ***/t/***, while the ***/ɨ/*** in A occurs after ***/z̧/***. As usual, especially in generative phonology, when sound environments are discussed, the long underscore '__' represents the position of the respective sound in question.

and sometimes to */t/* (as in these examples, see also the next section) in the other languages and apparently is an innovation of A. Furthermore, we found no occurrence of the sound sequence */zi̱/* in A (one can think easily of articulatory reasons for this fact).

A plausible explanation is, then, that SM and pTG maintained an original ***/i/* of pMATG in these and similar words while in A ***/i/* changed to */i/* in this environment, as */zi̱/* is not possible in A.[17] Therefore, this CS is no argument for any internal classification of Mawetí-Guaraní (in particular, no argument to group TG together with SM, opposed to A).

After elimination of this exception, it can be stated that vowels are indeed usually quite stable in Mawetí-Guaraní and do not serve as argument for an internal classification. Minor cases of not matching vowels in cognate words don't show any regular pattern and must be explained case by case.

There are, however, long vowels in Sateré-Mawé, but not in Awetí, nor are they reconstructed for pTG. As these long vowels in SM are not predictable by morphological or phonological rules, and as they in several instances correspond to long vowels in other Tupian languages (several have this contrast between long and short vowels), we propose here that pMATG also had long vowels in contrast to short vowels, and that A and the Tupí-Guaranian languages (probably already pTG) lost this contrast. This is indeed a shared change common to A and pTG and could be counted as an argument in favor of a closer sub-grouping of these two branches (i.e., configuration A in Table 1). However, loss of the contrast between long and short vowels is not so rare a change (if long vowels existed in proto-Tupí, as it seems, they were lost in various branches of Tupí), so the argument is rather weak as the change may have occurred independently on the way from pMATG to A and to pTG.

c. Other stable correspondences: stops, nasals, a tap, and glides

In most cases, the consonants are quite stable within Mawetí-Guaraní words, too. In Table 9 the most common correspondences are listed. Again, we will here not give special examples to illustrate these facts, since they are visible in the examples for other CSs, above and below.

---

[17] Note that the reconstructed hypothetical sounds or words of the proto-language pMATG are marked by two asterisks **, in order to differentiate these from the reconstructions of pTG, marked by one asterisk. For an alternate analysis for at least some of the cases discussed here, see the section 0.

Table 9 : Most common correspondences between stops, nasals and glides

pTG	Awetí	Sateré-Mawé
p	***p***	***p***
m	***m***	***m***
t	***t***	***t***
n	***n***	***n***
k	***k***	***k***
ŋ	***ŋ***	***ŋ***
ʔ	***ʔ***	***ʔ***
r	***r***	***r***
j	***j***	***j***
ß	***w***	***w***

The only CS which does not show identical elements in al three languages is that of the labio-velar glide ***w*** in A and SM, which often corresponds to a voiced labial fricative (or a bilabial approximant) ****ß*** reconstructed for pTG. However, although good reasons for this reconstruction for pTG exist, a contrast between a glide ****w*** and the approximant ****ß*** is a perceptually difficult contrast and may therefore be questioned. It there was no such contrast in pTG, the reconstructed ****ß*** would occupy the structural position of a glide (there is no other voiced fricative or approximant), analogous to the ****w*** in A and SM, and should itself also possibly better be represented as ****w***, making the apparent difference in the last line in Table 9 spurious. This makes the CS ***ß : w : w*** a possible candidate for differences generated by divergent analyses rather than by different phonological facts. Some of such cases are presented in the next section.

d. Differences generated by analysis

Observe the final segments in the cognate sets given in Table 10.

Table 10 : Examples for spurious different final segments

gloss	pTG	Awetí	Sateré-Mawé
snake	***moj***	***mõj***	***moi***
to finish	***paß***	***pap***	***pap***
to fall	***ʔar***	***ʔat***	***ʔat***

In the last two cases, pTG shows a different final consonant than A and SM. In the first case, it is SM which is deviant from the other two languages. These examples are representative; we can establish the CSs listed in Table 11.[18]

Table 11 : Correspondence sets for spurious different final segments

Environment	pTG	Awetí	Sateré-Mawé
V__.	***j***	***j***	***i***
__#	***ß***	***p***	***p***
__#	***r***	***t***	***t***

In the first case, the pronunciation of the two sounds ***j : i*** (for instance, in the three words for 'snake') may well be almost the same. The difference lies in the analysis of the last segment as a glide (consonantal segment), or as a vowel or part of a diphthong. This has virtually no impact on the reconstruction and does not serve as any argument for an internal classification of Mawetí-Guaraní.

The other two cases are similar. There is a phonetic rule of 'lenition' for word-final consonants in all three branches of Mawetí-Guaraní, which determines for instance that a bilabial consonant in that position will be pronounced as a ***[p̚]*** at the end of an utterance, but as an bilabial voiced fricative or even approximant ***[ß~β̞]*** before a following word beginning with a vowel. Similarly, the dental consonant in this position can be pronounced as ***[t̚]*** or as ***[r]*** in the same circumstances (A has lenition even for a final ***/k/***, pronounced as, e.g., ***[k̚]*** or ***[ɣ]***). This lenition rule has been reconstructed for pTG (see JENSEN, 1998; SCHLEICHER, 1998) and should be assumed for pMATG. Now, there may be reasons to reconstruct the phonemes ****/ß/*** and ****/r/*** for pTG, but as the

[18] The Symbols in the environment indications are: '#': word boundary, 'V': any vowel, and '.': syllable boundary. So, '__#' means 'end of word', and 'V__.' reads 'after vowel at the end of the syllable'.

allophonic rules are at least very similar in the three branches, the difference in word-final position appears not to be a substantial one, but a mere effect of different analyses.[19]

e. Differences generated by divergent groupings of allophones

There are regular correspondences that involve a stop in the case of SM and A, but the correspondent homorganic nasal for pTG, as indicated in Table 12 and exemplified in Table 13.

Table 12 : Correspondence sets involving nasals vs. homorganic stops

pTG	Awetí	Sateré-Mawé
****m***	***p***	***p***
****n***	***t***	***t***
****ŋ***	***k***	***k***

Table 13 : Examples for nasals vs. homorganic stops

gloss	**pTG**	**Awetí**	**Sateré-Mawé**
woman's child	****memir***	***mẽpit***	***mẽpit***
to listen	****enuß***	***ẽtup***	***nẽtup***
mortar	****iŋuʔa***	***ẽkuʔa***	***wẽkuʔa***

As can be seen in the examples in Table 13, the differing segments share all the same environment Ṽ__V: The vowel at the right of the segment is oral, but the vowel at the left is nasal – at least phonetically, in the case of pTG, due to the nasal spreading from right to left that is operative in Mawetí-Guaraní. Now, according to the allophonic rules of pTG, the phonetic variant of a nasal in this environment is a complex segment with a pos-oralized phase, voiced or possibly even voiceless. Also, in the case of Awetí and SM, oral stops in this environment have a pre-nasalized and optionally voiced phonetic variant. That is, e.g., ***/t/*** in A and SM as well as ***/n/*** in pTG have, in these environments, the same variants ***[nt]*** or ***[nd]***. Therefore, the phonetic variants of the examples in Table 13 given in

[19] In any case, the CSs are in complementary distribution to the 'regular' identity CSs, which by itself is sufficient to postulate only one proto-phoneme. If ****/ß/*** is maintained for pTG, the phonological assignment has to be analysed.

Table 14 are all the same or very similar. (For the final stops in the first and second rows, see the preceding subsection 0.)

Table 14 : Phonetic variants of examples for nasals vs. homorganic stops

gloss	**pTG**	**Awetí**	**Sateré-Mawé**
woman's child	****[mẽmbɨˀ]***	***[mẽmbɨˀ]***	***[mẽmbɨˀ]***
to listen	****[ẽndupˀ]***	***[ẽndupˀ]***	***[nẽndupˀ]***
mortar	****[ĩŋguʔa]***	***[ẽŋguʔa]***	***[wẽŋguʔa]***

Therefore, the differences on the phonological level disappear when we compare the phonetic realizations. What differs is not the sound shape itself, but the alignment of the allophones to phonemes, as summarized in Table 15, where '::' marks the contrast between the two phonemes, whose main allophones (in […]) are grouped between slashes '/…/'.

Table 15 : Distribution of allophones involving nasals vs. homorganic stops

proto-Tupí-Guaraní	**Awetí + Sateré-Mawé**
/[m] + [mb]/ :: /p/	***/m/ :: /[mb] + [p]/***
/[n] + [nd]/ :: /t/	***/n/ :: /[nd] + [t]/***
/[ŋ] + [ŋg]/ :: /k/	***/ŋ/ :: /[ŋg] + [k]/***

This regrouping by itself could be taken as an argument for a closer relationship between A and SM, if it could be demonstrated that the allophonic grouping is a shared innovation. However, comparison with other Tupian languages makes it more probable that A and SM preserve an older state of affairs while the current situation in most TG languages is the result of a common change that probably took place after pre-pTG split from the other two branches (in whatever order, if any). Therefore, the different phonological treatment of pre-nasalized stops or post-oralized nasals cannot be used as an argument for the internal sub-grouping of MATG at all.

6. First results of phonological (lexical) comparison 2: two vs. one reflexes

f. Tupí-Guaraní versus Awetí and Sateré-Mawé

In the following CSs given in Table 16, pTG diverges from A and SM.

Table 16 : Correspondence sets opposing pTG to A and SM

pTG	Awetí	S.-Mawé
****Ø***	***ʔ***	***ʔ***
****tʃ***	***t***	***t***
****ts***	***t***	***t***

Table 17 : Examples for pTG in opposition to A and SM

Gloss	pTG	Awetí	Mawé
to eat (itr., to have a meal)	****karu***	***karʔu***	***katʔu***
to take a bath	****atʃuk***	***atuk***	***atuk***
to go	****tso***	***to***	***to***

In the case of the first correspondence involving a glottal stop in A and SM, it seems obvious that pTG (or many single TG languages) lost this glottal stop that supposedly was present in pMATG and which is preserved by the other two languages. This seems to be a regular loss for pTG if the glottal stop occurs adjacent to another consonantal segment (cf. 'beatle': pTG ****maŋaŋa*** vs. A ***makaŋʔa***; 'jaguar': pTG ****jawar*** vs. A ***taʔwat*** etc.).[20] Again, as the deviant language seems to be innovative here, this does not contribute anything to the internal sub-grouping of Mawetí-Guaraní.

[20] SCHLEICHER reconstructs ****jaʾwar*** (SCHLEICHER, 1998: 24ff). The analysis of the glottal stop in Tupian languages is object of discussion. SCHLEICHER (ibid.), for one, interprets it as the manifestation of a (supra)segmental feature. This would explain why in Awetí lenition of ***/t/*** to ***[r]*** does take place not only before vowels, but, at least in some cases, also before the glottal stop (cf. ***karʔu***, above, with incorporation of ***kat*** 'thing' into the transitive verb stem ***ʔu***, 'to eat').

The same does not hold for the two other correspondences.[21] Here, it seems to be pTG that preserves older segments, while these in both, A and SM, underwent a change to ***/t/***. This can be counted as a possible argument to group A and SM closer together. However, as we will see below, a change of a sound to ***/t/*** is, in the case of Awetí, a very common phenomenon – Awetí has ***/t/*** as a reflex of many other sounds, especially complex ones. What is uncommon in the cases exemplified above is that Mawé undertook the same or an analogous change. In any case, the argument for a sub-grouping conforming configuration C (cf. Table 1) is rather weak.

g. Awetí versus Tupí-Guaraní and Sateré-Mawé

There are also some cases of regular correspondences that set Awetí apart from the other two branches, cf. the sets in Table 18 and the examples in Table 19.

Table 18 : Correspondence sets opposing A to pTG and SM

pTG	Awetí	Sateré-Mawé
n	***j***	***n***
r	***z̦***	***r***
t	***z̦***	***t***

Table 19 : Examples for A in opposition to pTG and SM

gloss	pTG	Awetí	Sateré-Mawé
(to be) fetid	****nem***	***jem***	***nem***
vulture	****uruβu***	***uz̦uwu***	***uruwu***
shoulder	****atiʔip***	***az̦iʔip***	***tiʔipi***

As for the first correspondence, it seems plausible that there occurred a regular change in A, in a nasal environment (i.e., when between two phonetically nasal vowels), from ***/n/*** (preserved in the other two branches) to ***/j/***, which is pronounced ***[ɲ]*** in this environment. Again, the deviant language (A) seems to

[21] It is again still in debate if these are one or two correspondences, as the phonological distinction between ***[tʃ]*** and ***[ts]*** in pTG is unclear. At least SCHLEICHER (1998) defends that these were in free or dialectal distribution.

have undergone a change, without any significance for the question of the internal sub-grouping.

Similar can be said about the other two cases, which both involve Awetí ***/z̪/***, a sound particular to Awetí to begin with. As shown, this retroflex voiced fricative sometimes corresponds to ***/r/*** in the other languages and sometimes to ***/t/***.

As there are several examples where a pTG ***/r/*** corresponds to ***/r/*** in A,[22] we start with the hypothesis that there were two different ***/r/***-like sounds in the proto-language, unless we discover some regular difference in the environment of these correspondences that would explain why sometimes A shows a reflex ***/r/*** and sometimes a reflex ***/z̪/***. A possible candidate for an additional proto-sound would be a palatalized ****/r^{j}/***, the articulation of which is somewhat closer to a retroflex sound than that of a simple ***/r/***. But this hypothesis has to be checked with more examples, especially from SM.

If there were indeed two different sounds in pMATG that have ***/r/*** reflexes in pTG and in SM, one of which developed into ***/z̪/*** in A and the other into ***/r/***, then the merger of both to ***/r/*** in pTG and in SM may be taken as an indication for a common history of these branches. If there was only one segment that developed for some yet unknown reasons into ***/z̪/*** and ***/r/*** in A, then this is most probably a development only in A and does not indicate a common development of the TG and the SM branches.

These remarks by and large carry over to the other CS, ***t : z̪ : t***. There are several cognates between pTG and A where both show a ***/t/*** in analogous environments, so for the time being we postulate an additional sound in the proto-language, possibly again a palatalized sound, ****/t^{j}/***.[23] This hypothesis would

---

[22] Examples are 'fish': pTG ****pira***, A ***pira?it***, SM ***piira***, or 'rasp, scrape': pTG and A ***karãj***. Unfortunately, there are very few SM cognate words (giving cognate triplets) that would show the SM reflexes of the critical sounds in question. The situation is complicated by the fact that in several cognates an ***/r/*** may be indeed a lenis variant of an underlying ***/t/*** at the end of a stem (as in ***karu*** vs. ***kar?u***, in Table 17, cf. footnote 20).

Another question is if ***/t/*** and ***/r/*** may have been variants of a single phoneme at some former stage, as the distribution of ***/r/*** is usually restricted to the medial position, and there is the relation between both sounds by the lenition rule in the morpheme-final position (in analyses for pTG, discussed in section 0).

[23] The only triplet with a correspondent SM word is 'corn': pTG ****aβati***, A ***awati***, SM ***awati***. This word is quite widespread even outside the Tupian languages and may well be a loan in

also account for the change of the vowel ***/i/*** to ***/i/*** in SM in the cognate set 'shoulder' in Table 19. On the other hand, it is difficult to imagine the protolanguage to distinguish betweeen all of these, ****/t/***, ****/ts/***, ****/tʃ/***, and ****/tʲ/***.

In any case, the indications of a closer relationship between TG and SM are weak, if existent at all.

h. Sateré-Mawé versus Awetí and Tupí-Guaraní

Finally, there are some correspondences that show one reflex for SM and one other for A and pTG. See the CSs in Table 20 and some examples in Table 21.

Table 20 : Correspondence sets opposing SM to A and pTG

pTG	Awetí	S.-Mawé
****k***	***k***	***ŋ***
****t***	***t***	***s***
****t***	***t***	***h***
****t***	***t***	Ø
****t***	***t***	***j***

---

one or several of the languages. More cognate words from SM are still missing. Good examples without SM are 'leg': pTG and A ***etima***, or 'moustache': pTG ***amotaβ*** and A ***apõtap*** (cf. the last section for a discussion of ***m : p*** and ***β : p***).

Indeed, many of the cognates that would exemplify the identity correspondence set ***t : t (: t)*** may be questioned when the ***/t/*** is in medial position (V__V), as they several times could be morphologically complex so that the ***/t/*** was originally in initial or final position. Take for instance the cognate pair 'swim': pTG ****itap***, A ***ɨtatap***, which possibly is a compound containing ***ɨ*** 'water' and an element ***tap*** ('to cross'?, reduplicated in A with following reduction of consonant cluster).

Table 21 : Examples for SM in opposition to A and pTG

gloss	**pTG**	**Awetí**	**Mawé**
wasp	****kap***	***kap***	***ŋap***
armadillo	****tatu***	***tatupep***	***sahu***
(to be) good	****katu***	***katu***	***kahu***
fire	****tata***	***taza***	***aria***
heel	****pita***	***pi apita***	***pia***
curassow	****mitũ***	***mituŋ***	***mijũ***

As for the correspondence of an initial stop to its homorganic nasal in SM, it seems possible that this is related to a regular alternation pattern that occurs in many languages, at least in the case of the bilabial segments ***/p/*** vs. ***/m/***, as in A in the word for 'hand': the inflected ('possessed') forms show the stem ***/po/***: ***ipo, epo, nãpo*** 'my, your, his/her hand' etc., but the 'absolute' (not-possessed) form is ***mo***. Similar changes occur here and there also with the dental and the velar consonants, as is here the case in SM. It is not possible yet to evaluate the question if these are rests of a general pattern that ceased to exist except for the bilabial consonants, or if these are analogous formations or generalisations. In either case, it would be a rather weak argument for grouping pTG and A together.

All the other examples have ***/t/*** for both, pTG and A, where SM shows consonants (or even no segment) that can be interpreted as the result of a process of 'weakening' which is also well known for reflexes of pTG ****/ts/*** and/or ****/tʃ/*** in several Tupí-Guaranian languages.

For lack of cognates in SM, it is yet unclear how many and which proto-consonants should be reconstructed in order to account for all these cases. Some of the correspondences can be clarified by looking at cognates from outside Mawetí-Guaraní where some (in particular, those with SM ***/s/*** and ***/h/***) correspond to a phoneme that for the time being is labelled *****/T/***. At least the last CS ***t : t : j*** could be the reflex of a proto-phoneme like *****/t^{j}/***, but then one has to explain why A doesn't show ***/zɻ/***.[24]

[24] Similar holds for he second consonant in the cognates for 'fire'. SM ***aria*** corrobates the hypothesis of an earlier proto-segment *****/r^{j}/***, corresponding to A ***/zɻ/***, but why does pTG not have the expected ****/r/***, then?

These CSs have some significance for the internal classification of Mawetí-Guaraní if it can be proven that pTG and A share a sound change, which seems possible at least in some cases.

7. First results of phonological (lexical) comparison 3: other cases

Finally, we will examine some other sound correspondences, especially those which show three different reflexes. Most of them concern again the dental consonants, and many have the glide ****/j/*** in pTG. See the CSs in Table 22 and some examples for cognates in Table 23.

Table 22 : Other correspondence sets

pTG	Awetí	S.-Mawé
****r***	***l***	***r/n***
****t***	***z***	***r***
****k^{j}/k***	***t***	***k***
****ts/tʃ***	***t***	***h***
****j***	***t***	***s***
****j***	***t***	***w***
****j***	***t***	***h***
****j***	***t***	Ø
****j***	***k***	***w***

Table 23 : Examples for other correspondences

gloss	**pTG**	**Awetí**	**Mawé**
(to be) bitter	****roß***	***lop***	***nop***
tortoise	****jaßoti***	***tawozɨ***	***woworɨ***
door	****oken***	***oten***	***oken***
eye	****etsa***	***e-ta***	***eha***
cotton	****aminiju***	***amatitu***	***amõkiusu***
pot, pan	****jaʔẽ***	***taʔẽ***	***waʔã***
to blow	****peju***	***petu***	***pehu***
salt	****jukɨr***	***tukɨt***	***ukɨt***
axe	****jɨ***	***kɨ***	***ɨwɨ hap***

It is a particularity of A to have a phoneme ***/l/***, even if marginally so, in opposition to ***/r/***; ***/l/*** is unknown in most TG languages and in SM, which usually show ***/r/*** in the analogous position.[25] As the distribution of ***/l/*** in A is apparently unpredictable, we assume that there was a phoneme ***/l/*** already in the proto-language pMATG.[26] Possibly, ***/l/*** was a rare phoneme already in the proto-language and maybe even earlier in proto-Tupí. Loss of ***/l/*** is a shared feature of SM and pTG but not a strong argument for a grouping of these two branches.

The second CS, ***t : z̨ : r*** is rather rare (a further example is the cognate set 'fire' in Table 21). It has a middle position between ***t : z̨ : t*** and ***r : z̨ : r***, see above in section 0. It is improbable that this set is based on a different sound in the proto-language; rather, we assume that for some reason there has been a change either in pTG or in SM.

---

[25] Word-initially SM sometimes has ***/n/*** instead of ***/r/***, which again may be a relic of a rule of word-initial nasaliation.

[26] Obviously, it is also possible that A acquired ***/l/***. An argument in favour of this possibility is the fact that the existence of ***/l/*** may be something like an areal feature: also among the Cariban languages, ***/l/*** occurs mainly or only in Bakairi, Kuikuro and Ikpeng, all located in the greater Xingú area. However, this origin of A ***/l/*** would require an identification of the circumstances that caused or allowed a change from ***/r/*** to ***/l/*** in A, which does not seem to be predictable.

Almost all following CSs, as many before, involve the phoneme ***/t/*** for Awetí. Indeed, it is a particularly remarkable feature of Awetí that several different sounds of the proto-language developed to ***/t/*** in this language, especially all sorts of complex sounds.

For instance the proto-sound for the CS ***k : t : k*** may have been a *****/k/*** – restricted to the position before front vowels. Note that in all cognates for the more common CS ***k : k : k*** the environment is **__V$_{[-anterior]}$**, that is, before a central or posterior vowel, and that in A, there are almost no occurrences of ***/k/*** followed by a front vowel ***/e/*** or ***/i/*** within a morpheme. It is probable that the pronunciation of *****/k/*** in this position was palatalized *****[k^{j}]***, which possibly developed via ****[c]*** to ***/t/*** in A. Today, ***/t/*** is pronounced strongly palatalized before ***/i/*** as *****[tʃ]*** in A (a similar rule holds for Portuguese in large areas of Brazil), and this may in many cases relate to an older (phonemically) palatal consonant.

The next set, ***ts : t : h***, resembles the CS ***t : t : h*** discussed above in section 0. Indeed, what can be affirmed is that most pTG ***/ts/*** and ***/tʃ/*** correspond to A ***/t/***; the few available cognates from SM show ***/t/***, ***/s/*** and ***/h/*** but are not sufficient to arrive at a conclusion about possible patterns and exceptions.

The same holds for the next CSs that all involve a correspondence between pTG ***/j/*** and A ***/t/***, while SM shows different reflexes including ***/s/***, ***/w/***, ***/h/*** and Ø. If it was not for ***/w/***, one could think that these correspondences are all a variation of the CS ***ts : t : h*** and the CSs discussed in section 0, with an additional alternation of *****/tʃ/*** or *****/ts/*** to ****/j/*** in pTG.

However, the CS ***j : t : w*** is a quite frequent one, especially word-initially (besides 'pot, pan', see also the cognate set 'tortoise' in Table 23 and 'mosquito' in Table 8). In order to account for the corresponding SM ***/w/***, we tend to reconstruct a hypothetical proto-sound *****/k^{w}/*** for pMAGT.[27] An obvious alternative candidate is *****/t^{w}/***. However, the scarce evidence from other Tupian languages shows that at least in some occasions the original phoneme would have been a velar. Cf. for instance 'moon': pTG ****jatsi***, A ***tati***, SM ***waati***, Gavião ***gat ti***.

[27] This proposal has to be checked, as there are other cognates showing ***/k^{w}/*** or its close reflexes (at least word-initially) in pTG and A (cognates from SM are missing), such as 'hole': pTG ***k^{w}ar*** and A ***kwat***.

If this hypothesis can be confirmed, the common change from a velar to a dental sound in pTG and in A would render a strong argument indeed for grouping these branches together.[28]

Once the fundamental character of the CS ***j : t : w*** is recognised, most examples for ***j : t :*** Ø (cf. the cognates meaning 'salt') appear to be related since a possible condition for dropping of the initial ***/w/*** in SM is the position before ***/u/***. Given this state of affairs, the examples for the CS ***j : t : h*** (cf. the cognate set 'blow', although not at the beginning of a word) can be associated with this development, showing an incomplete deletion of the expected ***/w/*** in SM, or they can be explained along the lines of the CS ***t : t : h***, see section 0.

The only CS showing ****/j/*** in pTG related to ***/t/*** in A that apparently cannot be related to CS ***j : t : w*** is, then, the CS ***j : t : s***, which has possibly only the cognate set 'cotton', given above, which is rather uncertain, at least for the first part(s).

The last CS in Table 22, ***j : k : w***, has some similarity with ***j : t : w***, except for the velar ***/k/*** in A. It is possible that it goes back to the same (velar?) phoneme in the proto-language and that the change to a dental stop in A has been blocked by some circumstances. We will not speculate any further on this matter, as there are only two examples: Besides the cognate set 'axe' in Table 23, the only other possible complete example available so far is 'cooked': pTG ****jiß***, A ***okip*** and SM ***wip***.

Some of the CSs presented in this subsection are supported only by very few cognate sets. This holds even more for the possible cognates in Table 24 which, for the time being, have to be considered as aberrant or even exceptions. Only more data, especially from SM, will help to clarify if borrowing is involved or what are the special circumstances that caused these isolated sound changes.

---

[28] A possible scenario would be pMATG *****/k^{w}/*** → proto-Awetí-TG *****/t^{w}/*** → pre-proto-TG *****/t^{j}/*** → pTG ****/j/***, with *****/t^{w}/*** being a sound of the common ancestor of Awetí and pTG but not SM.

Table 24 : Possible cognate sets demonstrating isolated sound correspondences

gloss	**pTG**	**Awetí**	**Mawé**
to sleep	****kʲer***	***tet***	***ket***
man's older brother	****ike?ir***	***iti?it***	***ike?et***
to enter / get in	****ekʲe / *ike***	***etse***	***weke***
to smell	****etun***	***etsun***	
to suck	****piter***	***pitsot***	
stone + shell	****ita + *itã***	***kita + kitã***	

The first example 'sleep' exemplifies a possible CS ***kʲ : t : k***, pointing at a hypothetical proto-sound *****/kʲ/***. However, the reconstruction for pTG is dubious (SCHLEICHER (1998), for one, reconstructs pTG 'sleep' ****ket*** instead of ****kʲer***), so the palatalisation may be just phonetic and the source for this CS would be simply a *****/k/***. In this case, its treatment would be analogous to those for 'door', above.

Related cases are those of 'man's older brother' and 'to enter', showing variations of ***k : t : k***, but with the special quality of showing different relics of the original palatalisation. In the case of 'man's older brother', A has a vowel change from ***/e/*** to ***/i/***, resulting in quite a different phonetic shape, ***[itʃi?it˺]***. Yet in the case of 'to enter', A has ***/ts/***, and there are different proposals for the reconstruction for pTG: RODRIGUES and DIETRICH (1997) have ****ekʲe*** while MELLO (2002) reconstructs ****ike***.

As has been demonstrated above (cf. Tables 16 and 22), pTG ****/ts/*** and/or ****/tʃ/*** usually correspond to A ***/t/***. Curiously, the opposite holds as well: the few examples of A ***/ts/*** with cognates in pTG usually show a correspondence to pTG ****/t/***, as exemplified by the words for 'to smell' and 'to suck' (SM does not show a cognate). In the latter case, the pTG ****/e/*** instead of A ***/o/*** could be a relic of some palatal element which also gave rise to A ***/ts/***.

Finally, there are a few cases where A has a ***/k/*** which is missing in pTG, usually at the beginning or end of a stem and adjacent to a high vowel, most often to ***/i/***. These cases need explanation, especially as they contrast with the

cognate set 'unripe': pTG ***kir***, A+SM ***kit***. A possible source would be *****/kʲ/***, but this does not explain the cases where the loss occurred word-finally.[29]

With these remarks we conclude the presentation of first results of our comparative research.

Conclusion

In this study we have examined the evidence for the exact genetic position of the Awetí language in the large Tupí family, especially evidence for an internal classification of the larger branch of Tupí called "Mawetí-Guaraní" which comprises the Tupí-Guaraní family, Awetí and Sateré-Mawé.

As it turns out, we did not find any clear example of an uncommon sound change which would have happened after the separation of the antecessor of one branch but before the split between the other two. There is some just probability that Awetí belongs somewhat closer to Tupí-Guaraní within Mawetí-Guaraní (configuration A in Table 1), but we did not find any conclusive evidence.

All we have are some weak indications the majority of which, however, point in this direction:

- a higher number of cognates found between Awetí and proto-Tupí-Guarani;
- lexicostatistic results (number of cognates in a 100-item-word-list proposed by Swadesh);
- loss of long vowels in Awetí and Tupí-Guaraní, but not in Sateré-Mawé;
- some sound changes suggest that in the development to Awetí and to proto-Tupí-Guaraní velar segments changes to dental segments (cf. the discussion of the correspondence set ***j : t : w***);
- possibly some of the correspondence sets given in Table 20.

We consider it to be too soon to conclude that there is a branch Awetí + Tupí-Guaraní of Mawetí-Guaraní, opposed to Sateré-Mawé, but *if* there is any grouping, this hypothesis is most promising.

---

[29] Examples for occurrences at the end are: 'blood': pTG ****uwi***, A ***uwik***; or '(to be/feel) cold': pTG ****roʔi***, A ***joʔik***, SM ***raʔak***. There is one example in intervocalic position: 'honey/bee': pTG ****eir***, A ***ekit***., with the same characteristic correspondence pTG ***/i/***, A ***/i/***, which could be covered by the hypothesis of an underlying *****/kʲ/***.

Bibliography

DIETRICH, WOLF. 1990. *More evidence for an internal classification of Tupí-Guaraní languages*. Indiana Beiheft 12. Berlin: Gebr. Mann.

DRUDE, SEBASTIAN. 2002. "Fala masculina e feminina em Awetí". In: *Línguas Indígenas Brasileiras Fonologia, Gramática e História. Atas do I Encontro Internacional do Grupo de Trabalho sobre Línguas Indígenas da ANPOLL*, ed. by Ana Suelly Arruda Cabral and Aryon Dall'Igna Rodrigues, p. 177-90. Belém: Editora UFPA.

FRANCHETTO, BRUNA and HECKENBERGER, MICHAEL J. (eds.) 2000. *Os Povos do Alto Xingu. História e Cultura*. Rio de Janeiro: Editora da UFRJ.

GALUCIO, ANA VILACY and GABAS JR., NILSON. 2002. "Evidências para o agrupamento genêtico Karo-Puroborá, Tronco Tupí". Paper presented at *17° Encontro nacional da ANPOLL*, Gramado.

HECKENBERGER, MICHAEL J. 2000. "Estrutura, história e transformação: a cultura xinguana na *longue douree*, 1000-2000 d.C." In: *Os Povos do Alto Xingu. História e Cultura*, ed. by Bruna Franchetto and Michael J. Heckenberger, p. 21-62. Rio de Janeiro: Editora da UFRJ.

JENSEN, CHERYL JOYCE S. 1998. "Comparative Tupí-Guaraní Morphosyntax". In: *Handbook of Amazonian Languages*, ed. by Desmond C. Derbyshire and Geoffrey K. Pullum, p. 487-618. Berlin & New York: Mouton de Gruyter.

LEMLE, MIRIAM. 1971. "Internal classification of the Tupí-Guaraní linguistic family". In: *Tupí Studies 1*, ed. by David Bendor-Samuel, p. 107-29. Norman: SIL.

MEIRA, SÉRGIO and DRUDE, SEBASTIAN. in prep. "A Preliminary Reconstruction of Proto-Mawetí-Guaraní Segmental Phonology".

MEIRA, SÉRGIO, DEMOLIN, DIDIER, DRUDE, SEBASTIAN, GABAS JR., NILSON, GALUCIO, ANA VILACY, LACERDA, MARIANA, MOORE, DENNY, PICANÇO, GESSIANE, RODRIGUES, CARMEN LÚCIA REIS and STORTO, LUCIANA. in prep. "Lexicostatistics of the Tupí stock".

MELLO, ANTÔNIO AUGUSTO SOUZA. 2002. "Evidências fonológicas e lexicais para o sub-agrupamento interno Tupí-Guaraní". In: *Línguas Indígenas Brasileiras Fonologia, Gramática e História. Atas do I Encontro Internacional do Grupo de Trabalho sobre Línguas Indígenas da ANPOLL*, ed. by Ana Suelly Arruda Cabral and Aryon Dall'Igna Rodrigues, p. 338-42. Belém: Editora UFPA.

RODRIGUES, ARYON DALL'IGNA. 1964. "A classificação do tronco lingüístico Tupí". In: *Revista de Antropologia*, 12. p. 99-104.

—. 1984/85. "Relações internas na família lingüística Tupí-Guaraní". In: *Revista de Antropologia*, 27/28.

—. 1985. "Evidence for Tupí-Carib relationships". In: *South American Indian languages. Retrospect and prospect*, ed. by Harriet E. Mannelis Klein and Louisa Stark, p. 371-404. Austin: University of Texas Press.

RODRIGUES, ARYON DALL'IGNA and DIETRICH, WOLF. 1997. "On the linguistic relationship between Mawé and Tupí-Guaraní". In: *Diachronica*, XIV. p. 265-304.

SCHLEICHER, CHARLES OWEN. 1998. *Comparative and Internal Reconstruction of the Tupi-Guarani Langauge family*.

SOUZA, MARCELA STOCKLER COELHO DE. 2000. "Virando gente: notas a uma história awetí". In: *Os Povos do Alto Xingu. História e Cultura*, ed. by Bruna Franchetto and Michael J. Heckenberger, p. 360-402. Rio de Janeiro: Editora da UFRJ.

STEINEN, KARL VON DEN. 1894. *Unter den Naturvölkern Zentral-Brasiliens*. Berlin: Dietrich Reimer.