

FERNANDO O. DE CARVALHO

Museu Nacional, Universidade Federal do Rio de Janeiro (MN/UFRJ), Rio de Janeiro

DFG Center for Advanced Studies "Words, Bones, Genes, Tools", Eberhard Karls Universität Tübingen

fernaoorphao@mn.ufrj.br

ORCID: 0000-0002-2115-7416

AN IMPROVED ETYMOLOGY FOR THE TUPI-GUARANI ETHNONYM *KAGWAHIVA*¹

Keywords: etymology, Tupi-Guarani, ethnonyms, Kagwahiva

Abstract

The present contribution challenges the traditional etymology of the well-known ethnonym *kagwahiva*, which goes back to the first decades of the 20th century. It is shown that the hypothesized etymological association with the reflexes of the Proto-Tupi-Guarani **kap/*kaβ-* 'wasp' is formally untenable. An alternative proposal is presented, supported by argumentation at phonological, morphological, semantic and syntactic levels, and based on the identification of clear, yet so far, unacknowledged cognates in Old Tupi, Old Guarani and other languages within the family. A PTG etymon **-kawaiɸ/*kawaiβ-* 'to be aggressive, prone to violence' is tentatively proposed, and further etymologization is advanced on the grounds of formal and semantic associations to **-kaʔu* 'to drink (alcoholic beverage)'.

1. Introduction

The etymological investigation of ethnonyms has a somewhat irresistible power, in part due to the considerable potential it harbours for casting light on the (pre-) history of human groups (see e.g. Clauson 1963; Bačić 1987; Kempf 2011/2012).

¹ I am grateful to Kamil Stachowski and the two anonymous reviewers for their comments and careful reading of this article. Their observations, comments, and questions helped to improve this paper in many ways. It goes without saying that any remaining shortcomings are my own responsibility.

More often than not, however, rather than providing independent evidence about societies from the past, it depends, if undertaken correctly, on knowledge concerning the extra-linguistic world of these communities. In any case, the etymological study of ethnonyms, as a specialized branch within onomastics, is usually expected to face challenges not present in the etymology of “common words” (see e.g. Lass 1973; Kempf 2011/2012); besides which, caution is especially needed when trying to base conclusions about non-linguistic (pre-)history on the results of such studies (see e.g. Turin 2002). For indigenous South America, other than the beliefs of amateurs and intellectuals from early colonial times, the issues around the uses and origins of ethnonyms have been of perennial interest to anthropologists (e.g. Erikson 2004), although few works on this subject by professional linguists appear to have been published to date (see e.g. Cerrón-Palomino 2008; Ramirez 2010; Carvalho 2019).

The aim of this paper is to discuss the etymology of the well-known ethnonym *kagwahiva*. Although variants of this word have been employed in non-Tupi-Guarani groups of the south-western Amazonian region (see e.g. Menéndez 1989; Ramirez 2010), its use as a reference to certain Tupi-Guarani groups – and, in particular, as the autonym of specific sub-groups – constitutes the focus of the present paper. I will first show that the traditional etymology which relates this form to the reflexes of the Proto-Tupi-Guarani etymon **kap/*kaβ-* ‘wasp’ is untenable. After a discussion of the earliest attestations of this ethnonym (section 2), the formal problems regarding the currently accepted etymology will be addressed in section 3. An improved etymology is advanced in section 4, based on the identification of cognates of this ethnonym in the classical languages within the family (Old Tupi and Old Guarani), and also in some modern languages (Kamayurá, Wajãpi). It is determined that the ethnonym most probably originates in a (late) Proto-Tupi-Guarani form **kawaiɪ/*kawaiβ-* ‘to be aggressive, prone to violence’. A persistent problem in Proto-Tupi-Guarani lexical reconstruction stemming from the work of Lemle (1971) is also discussed and resolved. This, in turn, allows an explanation of the unexpected intervocalic *-h-* in *kagwahiva* – whose problematic character is unfortunately overlooked entirely in the existing literature – via contamination processes. On somewhat more tentative grounds, section 4.1 explores the hypothesis that **kawaiɪ/*kawaiβ-* can be further etymologized as stemming from **-kaɽu-aip* ‘to drink in excess’, thus relating the state of violence or unrest denoted by the term to outbursts of aggression induced by the consumption of alcoholic beverages. Finally, section 5 offers a general summary of the arguments and evidence presented.

2. Attestation

The following entry is found in the Betts (2012) pan-dialectal (yet mostly Parintintin) dictionary of *Kagwahiva*:

Kagwahiva s. used as a general term for Indians, especially the Parintintin and Tenharim, but which term excludes the Mura Pirahã Indians who were enemies of the

Parintintin and the Tenharim. Those groups who call themselves by this term are Parintintin, Tenharim, Karipuna of Rondônia, Juma, Uru-eu-wau-wau, and Amundava. (Betts 2012: 129)

The noun in question functions as an ethnonym, a proper name used as an autonym by the “Kagwahiva groups”, and, as such, it stands in direct opposition to terms employed for the Other (see also Kracke 1978: 16–17; Menéndez 1989: 69–71). The word-form reported by Betts (2012) agrees, in all relevant formal and semantic aspects, with the early form of the autonym reported by Nimuendajú (1924) in his classic ethnographic text on the Parintintin:

A tribo de índios vulgarmente conhecida por “Parintintin”, no Rio Madeira, em sua própria língua se denomina Kawahíb ou Kawahíwa, quando este nome ainda é seguido por um sufixo, uma posposição ou um adjetivo. (Nimuendajú 1924: 201)

‘The Indian tribe popularly known as the “Parintintin”, from the Madeira River, in their own language call themselves Kawahíb, or Kawahíwa when this name is followed by a suffix, a postposition or an adjective.’²

On the basis of both published and unpublished sources on the phonetics and phonology of *Kagwahiva* (notably: Pease and Betts 1971; Pease 1977; Pease and Betts 1991; Sampaio 1998; Betts 2012), the form ⟨Kagwahiva⟩, recorded by Betts (2012) in the practical orthography devised by SIL linguists for the language, can be straightforwardly associated with the phonetic form [kag^wa’hiβa], in turn phonologically analyzable as /kag^wa’hiβa/. I have opted in this text for the adoption of the orthographic variant *kagwahiva* when referring to the word under investigation, whether it is used as a glottonym or as an autonym.³

The forms ⟨Cabahibas⟩ and ⟨Cabahybas⟩, which can be observed on an anonymous 17th century map as the ethnonym of certain groups inhabiting the Upper Tapajos region (see Nimuendajú 1924: 205; Menéndez 1989), must be considered, for etymological purposes, to be distinct from *kagwahiva*.⁴ Even if ultimately related to the autonym *kagwahiva*, these are forms that have a history of their own, being the end-result of a long and obscure chain of transmission and adaptation that started with indigenous individuals and local non-indigenous middle-men until they were recorded in the existing documents. Their tortuous, and most likely unknowable, history may, in fact, include intentional deformations of the forms by the recorders, motivated by their own ideas on the etymologies of these names, which only adds to the complexity of the task of inquiring into their origins. That this is so is suggested, first, by the isolated

² Unless otherwise indicated, all translations are my own.

³ The difference between the use and mention of the ethnonym should be clear to the reader but, in case it is not, certain clarifications are presented: *kagwahiva*, in italics, without capitalization, is the convention used when mentioning the word; ‘Kagwahiva’, with initial capitalization, indicates the ethnonym is being used as a reference to the relevant group or language. Finally, angled brackets (as in ⟨Cabahiba⟩) will be used when specific attention is directed to aspects of a source’s orthography, or for the first mention of a given wordform.

⁴ A third form, ⟨Cauahipe⟩, was also recorded, this time by José da Silva Guimarães, after a meeting with Apiaká individuals who were visiting the city of Cuiabá (see again Nimuendajú 1924).

attestation of a medial ⟨b⟩ ([b] or [β]) that are only noted in these secondary records. This contrasts with the consistent recording of either ⟨w⟩ or ⟨gw⟩ in the more reliable, modern sources (⟨Cabahiba⟩ vs. ⟨Kagwahiva⟩). As will be seen below, the nature of these segments is critical in suggesting the direction of the most likely etymology for the noun *kagwahiva*. Second, and carrying perhaps more weight, the intentional deformation of a form that ultimately relates to *kagwahiva*, based on etymological assumptions, is in fact attested in Martius' rendering of ⟨Cahahybas⟩ (Martius 1867: 202). The deformation created by the exchange between the medial ⟨b⟩ and ⟨h⟩, which makes this form unique, probably stems from Martius' own hypothesis that the term might reflect a complex formation, including *kaʔa* 'woods' (see Nimuendajú 1924: 201). Finally, the very presence of the Portuguese plural suffix in ⟨Cabahibas⟩ and ⟨Cahahybas⟩ indicates their independent status as Portuguese words, thus capable of dissociation from the autonym *kagwahiva* employed by the relevant Tupi-Guarani groups.⁵

We are thus on much safer ground by restricting our attention to the native formation *kagwahiva* as recorded by professional linguists – or by linguistically trained observers – in modern times. One should not exclude, of course, the possibility that a formal development based on the speakers' perceived associations between etymologically independent forms (i.e. folk etymology) has affected this form too. In fact, we will have the opportunity to determine that one such non-*lautgesetzlich* development seems to have taken place in the case of *kagwahiva* (see section 4.1).

3. *Kagwahiva*: The traditional etymology and its shortcomings

The best-known proposal regarding the origin of this ethnonym is that of Nimuendajú (1924), which was later repeated, albeit in abbreviated form, in Nimuendajú's chapter on the *Kagwahiva* for the *Handbook of South American Indians* (see Nimuendajú 1948: 284). The original 1924 formulation, following on from the quote given above, is now presented in full:

A tribu de índios vulgarmente conhecida por “Parintintin”, no Rio Madeira, em sua própria língua se denomina *Kawahib* ou *Kawahíwa* quando este nome ainda é seguido por um sufixo, uma posposição ou um adjetivo. Não tem este nome a significação de “homens da matta”, como Martius explica (CM II. 3), mas é composto de *kab*, *káwa* = vespa + *ahíb* (=?), e designa uma pequena qualidade de vespas sociaes, de côr avermelhada e muito irritáveis que também entre os moradores do Baixo Amazonas é conhecida como “cauahiba”. (Nimuendajú 1924: 201)

‘The Indian tribe, popularly known as the “Parintintin”, from the Madeira River, in their own language call themselves *Kawahib* or *Kawahíwa*, when this name is followed by a suffix, a postposition or an adjective. This name does not bear the meaning of “men of the woods”, as explained by Martius (CM II. 3), but is composed instead of

⁵ Kracke (1978: 253) seems to agree with this overall conclusion, noting that “Cabahibas” may be conveniently applied only to the Upper Tapajos groups of which the modern *Kagwahiva* are the presumed descendants.

kab, *káwa* = wasp + *ahib* (=?), and designates a small kind of social wasp, of reddish color and very aggressive, which is known as “cauahiba” by the inhabitants of the Lower Amazon as well.’

As noted, an initial problem with this proposal is the absence of any account of the residue *-ahib*, which remains once the presumed base *kawa-* is identified. Later commentators have either remained uncertain concerning the plausibility of this proposed etymology (e.g. Ramirez 2010), or seem to have tacitly accepted it (e.g. Menéndez 1989: 69–70). More difficult to classify is the position of Kracke (1978), who states that *kagwahiv* was “incorrectly recorded” as “Cabahiba”, the “Bee-people” (Kracke 1978: 253). Aside from the obvious imprecisions (the failure to distinguish “bee” and “wasp”), and from the misleading suggestion that Nimuendajú could have misreported *kagwahiva* as <Cabahiba>, one is left wondering whether the author fails to accept the form or the meaning, or both. Be that as it may, a recently published study, Aguilar (2017), again addresses the Nimuendajú etymology, but adds what she considers an improved formulation of the proposal (Aguilar 2017: 149).

[...] uma composição de *-kaβ* ‘caba’ e *-ahib* ‘ruim, braba, forte’, resultando em *Kawahib* ou *Kawahib-a*, em que o tema é flexionado pelo caso argumentativo: *Kawahibα* ou *Kawahíwa*. (Aguilar 2017: 149)

[...] a compound of *-kaβ* ‘wasp’ and *-ahib* ‘bad, angry, strong’, resulting in *Kawahib* or *Kawahib-a*, where the stem is inflected by the argumentative case: *Kawahibα* or *Kawahíwa*.

The etymology advanced by Aguilar (2017) seeks to rectify the most obvious shortcoming in the Nimuendajú etymology by ascribing a morphological analysis to the residue *-ahib*, – in this instance *ahib*, whose meaning is given as ‘bad, strong, mean’. As discussed below, however, this traditional etymology cannot be accepted in either of its formulations.

The most obvious shortcoming of the etymological association between *kagwahiva* and <kav-a> ‘wasp’ (Betts 2012: 135) is the mismatch between the labialized voiced velar stop *-gw-* attested in the ethnonym and the fricative *β* (= <v>) found in its presumed base. *Kagwahiva* is one of the TG languages retaining the PTG opposition between **w* and *β* (see Lemle 1971; Schleicher 1998). The *Kagwahiva* reflex of the PTG **w* is usually represented as <gw> in sources focusing on the language, while the reflex of the PTG **β* is transcribed as <v>. The phonetic values for these orthographic symbols are, according to Betts (2012: 2–3), as follows: <gw> = [gʷ], <v> = [β]. The former is described as a ‘labialized voiced velar stop’, the latter as a ‘voiced bilabial fricative’. Table 1 below illustrates the retention of the **w* - **β* contrast in *Kagwahiva* through a comparison of the *Kagwahiva* forms with their cognates in Old Tupi, an early attested TG language, and the PTG etyma underlying each of the equations (Old Tupi data from Drummond (1952, I: 61, 70; II: 56, 67, 140), PTG forms from Schleicher (1998: 328, 337, 351) and Mello (2000: 178)).⁶

⁶ The interpretation of the relevant Old Tupi graphemes as *w* = <gw> and *β* = is supported by Barbosa (1956) and Rodrigues (1958), to which the reader is referred. Additional comments on Old Tupi <gw> are nonetheless offered below in section 4.

	Kagwahiva	Old Tupi	PTG
‘hair’	<i>-ʔaβa</i>	⟨Aba⟩	*-ʔáβ
‘sky’	<i>ibaga</i>	⟨Jgbaca⟩	*iβák
‘to fly’	<i>-βeβe</i>	⟨Abebe⟩	*-βeβé
‘jaguar’	<i>ɕʒagwara</i>	⟨Iaguara⟩	*jaʔwár
‘bird’	<i>gwira</i>	⟨Guirâ⟩	*wirá
‘slow’	<i>mbegwe</i>	⟨Begue⟩	*mewé

Table 1: PTG contrast between *w and *β retained in Kagwahiva

Among the forms that retain a voiced bilabial fricative is the Kagwahiva form for ‘wasp’. Since this PTG etymon is reconstructed neither by Lemle (1971) nor by Schleicher (1998, though see *kaβ in Mello 2000: 170), table 2 offers comparative evidence from a representative sample of eleven TG languages, thus providing verification for the reconstruction of the PTG *kap/*kaβ- ‘wasp’.⁷

Language	Form	Meaning
PTG	*-kap/*kaβ	‘wasp’
Old Tupi	⟨Caba⟩	‘bespas’ [‘wasps’]
Tenetehára	kaw	‘vespa, caba’ [‘wasp’]
Tocantins Asurini	kap, káwa	‘vespa, caba’ [‘wasp’]
Kamayurá	kap	‘vespa’ [‘wasp’]
Wajãpi	kaa	‘guêpe, mouche’ [‘wasp, fly’]
Kayabi	kap, kawa	‘marimbondo, caba’ [‘hornet, wasp’]
Guarayu	⟨Cau⟩	‘Wespe’ [‘wasp’]
Old Guarani	⟨cab⟩	‘abispá’ [‘wasp’]
Kagwahiva	kava	‘wasp’ [‘wasp’]
Avañe’ẽ	Cava, ca	‘avispa’ [‘wasp’]
Mbyá	kavy	‘vespa’ [‘wasp’]

Table 2: Cognate set supporting the reconstruction of *kap/*kaβ-

In the table above the cognates are given as they appear in the sources, with bold font highlighting the root in each word form (that is, the actual reflex of the PTG *-kap/*kaβ). The residue, usually a final -a or, in some Guaranian lects, -i ⟨y⟩, stands for a reflex of the nominal function marker *-a (see Jensen 1998), which depending on the language and the specific form, may have come to be incorporated as part of the root. Without going into the details of PTG morphophonology, the language can be reconstructed with an alternation between fortes *-p, *-t, *-k and, respectively, lenes *-β, *-r, *-ɣ root-final consonants, with the latter alternants occurring

⁷ Sources for the data: Old Tupi (VLB, I, 55), Tenetehára (Boudin 1978: 101), Tocantins Asurini (Cabral and Rodrigues 2003: 97), Kamayurá (Seki 2000: 372), Wajãpi (Grenand 1989: 206), Kayabi (H.E. Weiss 2005: 48), Guarayu (Hoeller 1932: 50), Old Guarani (Restivo 1893), Kagwahiva (Betts 2012: 135), Avañe’ẽ (Jover Peralta and Osuna 1950: 217), Mbyá (Dooley 1998: 193).

whenever a heteromorphemic (usually, suffixal) vowel follows (Jensen 1998, 1999; Schleicher 1998).⁸ This is most transparently seen above in the Kayabí cognates, with *kap* [kapʰ] appearing as *kaw-* [kaw-] when followed by the nominal function suffix *-a*. This is, therefore, the reason for reconstructing the alternation **kap/*kaβ-* ‘wasp’.

The existing sources on *Kagwahiva* make it clear that the language retains *β*, the orthographic ⟨v⟩ as the reflex of the PTG **β*, even in those cases where it originally alternated with the fortis **-p*, and is in clear opposition to *gw*, the orthographic ⟨gw⟩, the reflex of the PTG **w*. Regarding the form for ‘wasp’, Betts (2012) recognizes both a noun ⟨kava⟩ ‘wasp’ (with a root ⟨kav-⟩ followed by the nominal function suffix), and a number of derivatives naming specific kinds of wasps: ⟨kavaʼmbauhua⟩ ‘wasp that lives in trees’, ⟨kavavajuva⟩ ‘wasp that lives in trees’, ⟨kaviagwara⟩ ‘green wasp of the beach’, ⟨kaviakohoa⟩ ‘wasp of the beach’ (Betts 2012: 135–136). There are, therefore, compelling reasons not to accept the analysis of *kagwahiva* as yet another derivative of *kaβ-* ‘wasp’ in *Kagwahiva*, unless one is prepared to acknowledge that speakers have ignored a consonantal opposition, freely switching between *w* and *β*, and that they did so in this derivative formation alone.

4. An improved etymological proposal

The etymology proposed centres on the fundamental fact that the wordform *kagwahiva* has plausible cognates elsewhere in the family and, most notably, within the classical languages of the family: Old Tupi and Old Guarani. The Old Tupi cognate can be harnessed from the main lexical source on the language, the (probable) 16th century *Vocabulário na Língua Brasílica* (VLB):

⟨**kaguaíba**⟩ ‘pessoa exaltada, feroz, arrebatada, brava’ [‘exalted, ferocious, rampant, angry person’] (VLB, I, 45).

⟨**Xe kaguaíb**⟩ ‘Eu sou exaltado’ [‘I am exalted’] (VLB, I, 42).

The former form, analyzable as *kawaiβ-a*, with the nominal function marker *-a*, appears typically in nominal functions. Without this suffix, that is, as *kawaiβ*, it functions as a stative predicate, as shown by the occurrence of the Set II first person singular prefix ⟨Xe⟩ (more on this below).

On the formal aspects of the proposed equation, it should be noted that Old Tupi ⟨gu⟩ (sometimes ⟨go⟩) was a common allograph for representing [w] particularly in the

⁸ A detailed reconstruction of PTG consonantal morphophonology lies outside of the focus of this paper, but is currently being undertaken by the author. Most published overviews, such as those just cited, do little more than project back to PTG a situation attested in Old Tupi. As a consequence, there is hardly any mention of the fact that this lenition alternation probably targeted **-k* as well, a fact not noted since this consonant is not subjected to lenition in Old Tupi, the implicit stand-in for an appropriate reconstruction of PTG. Moreover, the pre-pausal/word-final alternant is routinely reconstructed with the lenes variants, again reflecting a situation that holds for Old Tupi. A comprehensive evaluation of the comparative evidence suggests, however, that fortes variants were found instead in the pre-pausal position, as is the case in languages such as Kamayurá, Kayabí and Xingu Asurini.

Portuguese Jesuit sources, such as Anchieta (1595) and Figueira (1687) (see Rodrigues 1958). Though the issue cannot be addressed within the limits of the present paper, it is likely that Old Tupi ⟨gu, go⟩, often transliterated as [w] also had a velar constriction of some kind.⁹ Of greater importance is the fact, noted before in table 1, that Old Tupi, like Kagwahiva, is one of the phonologically conservative languages retaining the PTG contrast between *w and *β. Thus, the external comparison of Kagwahiva *kagwahiva* and the Old Tupi *kawaiβ-* offers yet another reason for doubting the internal association with *kaβ-* that is central to the traditional Nimuendajú etymology.

On the distributional side, many early studies on comparative TG linguistics accepted sets with cognates in Old Tupi and Kagwahiva alone as evidence for the reconstruction of a PTG etymon (see e.g. sets 102, 133, 146, 172 in Lemle 1971), no doubt based on an implicit link to geographic factors, which would rule out the explanation of such a match as a secondary innovation in one of these languages with later diffusion to the other. Although in general this reasoning is compelling, there is no need for such a limitation to be followed in the present case. In Kamayurá the noun ⟨kawa'ip⟩ ([kawa'ipʔ]) is used as a reference to members of groups outside the Upper Xingu cultural area (see Silva 1981). In the overall ideological profile of the Xinguano culture, indigenous groups not belonging to the Upper Xingu acculturation zone are usually described as having a tendency towards all kinds of violent and unpredictable behaviour (see many of the contributions in the collection of Franchetto and Heckenberger 2001), which is entirely in line with the meaning recorded for Old Tupi. Along the same lines, the Kayabí use *kawaiip* [kawa'ipʔ] as a name for the Rikbáktsa (Weiss 2005: 50). In this instance too negative connotations are associated with a term for outsiders: The Kayabí consider the Rikbáktsa as inferior, often commenting disrespectfully upon and ridiculing their customs (see Grünberg 2004: 179).¹⁰ The inclusion of the Kayabí and Kamayurá cognates make the proposal of a PTG provenance for the Old Tupi *kawaiβ-* : Kagwahiva *kagwahiva* equation distributionally more persuasive. I consider it plausible that a PTG etymon **-kawaiip*/**kawaiβ-* 'to be aggressive, prone to violence' can be reconstructed on the basis of these (and other) cognates, and the remainder of this section will provide further justification for this claim.

On the semantic side of the proposed equations, the comparisons above suggest a meaning field encompassing the expression of violent behaviour, possibly

⁹ Remarks on the comparative status of the velar gesture in the approximant *w* have been vague and rather randomly located in the literature on Tupi-Guarani languages, a situation that often leads to contradictory or inconsistent claims. Jensen (1999), in her overview of the Tupi-Guarani language family simply states that the realization of **w* is [gw] in some languages of this group (C. Jensen 1999: 135). An earlier overview by the same author is more precise, claiming that the PTG **w* > *gw* in Kaiowá and that **w* > *gu* in Old Guarani and Mbyá Guarani (Jensen 1998: 605). Dietrich (1990: 20), in a widely cited study on TG internal classification, treats the existence of a [gw] allophone for the glide *w* as a primitive property of PTG, one that was lost in certain northern or Amazonian languages, except for Kagwahiva (Dietrich 1990: 20). Schleicher (1998) briefly discusses the correspondence between *w* in most TG languages and *ɣw* in Guarani Mbyá, Kaiowá and Kagwahiva, raising the issue of how to determine a correct reconstruction of the phonetic content of the relevant PTG segment (Schleicher 1998: 15).

¹⁰ The Rikbáktsa (also: Erikpatsa, Canoeiros) are speakers of a Macro-Jê language.

understood as a specific feature of outgroup members or hostile foreigners in general. Before addressing how this can be reconciled with the meaning of *kagwahiva*, and before advancing a specific PTG etymon underlying this set of cognates, I will discuss in greater detail another example, one with potentially significant implications for further etymologization. The Old Guarani cognate of the forms discussed above is introduced in Montoya's *Tesoro* as follows:

Caguaí: 'aporrear, maltratar el borracho, o el que no lo es' ['hit, mistreat, a drunkard, or someone who isn't one'], **Acaguaí hecê**: 'Maltratele' ['mistreat him/her'], **Ambocaguaí**: 'hazer que le aporree' ['to make (someone) hit him/her'] (Montoya 1639: 85v)

The first observation concerns the meaning of the Old Guarani form and its syntactic implications. The Spanish verb "aporrear", given as a translation of the Old Guarani <Caguaí> (= *kawai*), means 'to beat or hit with a club or wooden stick'. This is evident from Restivo's later *Lexicon Hispano-Guaranicum*, based to a large extent on Montoya's *Tesoro*, where a synonymous expression is found in <aynupa ýbira pipe>, which literally means "I hit him/her/it with a piece of wood". Next, consider that in comparison to the stative meaning reported for Old Tupi, the Old Guarani cognate has a more active meaning, one that involves carrying out an action that could be seen as typical of transitive verbs, as it seems to involve an affected participant as well. Nevertheless, the existing evidence suggests that the verb is in fact intransitive, as demonstrated by the following evidence: first, an affected participant must be introduced with a postpositional marker; second, the root *kawai* can be causativized with the use of the causative prefix *mo-* which is used with intransitives only; and, third, by the fact that the third person object prefix *i-* does not occur with these verbal constructions (see Jensen 1998, 1999). Relevant examples are: <acaguay hece> 'le aporree maltratandole mucho' (Restivo 1893: 90) and <Ambocaguaí> 'hacer que le aporree' (Montoya 1639: 85v), which are adapted and glossed below:

- (1) *a-* *kawai* *h-* *ese*
 1SG.I -beat.up 3- POST
 'I beat him/her up'
- (2) *a-* *mo-* *kawai*
 1SG.I CAUS -beat.up
 'I make him/her beat up'

As to their morphosyntactic behaviour, those familiar with the overall grammatical profile of the conservative TG languages may have noticed that while the Old Tupi *-kawaiβ* is a stative intransitive, evidenced by the use of the Set II first person singular marker *fe-*, its Old Guarani cognate is an active intransitive, occurring in the above examples with the Set I marker for the same person-number combination, *a-*.¹¹ As noted by Barbosa (1956), in Old Tupi, it is a property of many active

¹¹ For the difference between the two sets of person markers found throughout the TG language family, see the discussion in Jensen (1998, 1999).

intransitive verbs that they can be used either with Absolutive/Set II person indexes, or, instead, with the Active/Set I markers, with an aspectual distinction following from this choice. This fact was already highlighted in the first description of a Tupi-Guarani language, Anchieta's *Arte*, published in 1595. The following contrasts in the use of the two person-indexing sets are among those listed by Anchieta (as they appear in the original, with the person markers highlighted in bold font): <Anheêng> 'falo' ['I speak'] vs. <Xenheêng> 'sey, ou posso falar' ['I can speak']; <Aitâb> 'nado actu' ['I swim'] vs. <Xeitâb> 'sey nadar' ['I know how to swim']; <Ambaêcuâb> 'sey actu algũa cousa' ['I know something'] vs. <Xembaêcuâb> 'sou entendido' ['I am wise, knowledgeable'] (Anchieta 1595: 51–51v). This property, usually described as being diagnostic of a 'fluid-S system' in the literature on the typology of morphosyntactic alignment, has been commented upon specifically in relation to Guarani, as in the following passage in Mithun (1991):

Guarani speakers do have some choices; a few intransitive stems can appear with either case. The differences in their meanings confirm the semantic basis of the case distinction. The verb *karú* means 'to have lunch or supper, to dine' with the first case but 'to be a glutton' with the second. The verb *kaʔu* means 'to get drunk' with the first case but 'to be a drunkard, to be drunk' with the second [...] This does not imply, of course, that speakers must select a case every time they use a verb. Presumably they usually select a lexicalized case-plus-stem combination as a unit. (Mithun 1991: 513–514)

Mithun's concept that particular prefix-stem combinations¹² could be lexicalized is relevant here because it suggests a single etymon **-kawaip* can be reconstructed with different lexicalized choices of person-markers occurring in the two languages, namely Old Tupi and Old Guarani. Thus, whatever the meaning of the etymon itself, it is likely that Aktionsart-related alternations of the kind seen above for Old Tupi were in place, with **ʃe-kawaip* and **a-kawaip*, for a first person singular S argument, yielding, after semantic shifts, the reflexes attested, respectively, in Old Tupi and Old Guarani.

After accounting for the distinct morphosyntax of the Old Tupi and Old Guarani cognates by reference to the 'fluid-S' nature of (some) TG intransitive predicates, we need to understand how nominal cognates – such as those noted above for Kamayurá and Kayabí – can also be included as *bona fide* reflexes of a single etymon. This, in fact, is not of great significance: throughout the TG language family, intransitive verbs can be employed without further morphological elaboration as adnominal attributive modifiers. In Old Tupi, for instance, the verb <bebe> 'to fly', can be used to derive the name for 'angel' as a new coinage: <apiâbebe>, lit. 'flying-man' (see Drummond 1952: I, 36). The same is attested in the modern languages, such as Avañe'ẽ, where the intransitive verb *-karu* (*a-karu* 'yo como' ['I eat']), occurs as a simple attributive modifier in *mitã karu* 'niño comilón' ['gluttonous child'] (Krivoshin

¹² In the paper in question, Mithun refers to the two sets of Guarani person-markers, here called Set I and Set II markers in line with TG comparative linguistics, as being different "cases".

de Canese 1983: 47, 78). The nominal reflexes of a PTG etymon would derive from simple nominal ellipsis in phrases of the kind **aβa kawai-p-a* ‘aggressive person’ > **kawai-p-a* ‘aggressive (person)’.¹³

Table 3 below sums up the comparisons so far, with source transcriptions uniformly adapted to IPA transcriptions, and the meanings reported in each replaced by a convenient yet selective English summary of their respective meanings. I have also added two other cognates: the Wajāpi ⟨kaʔuai⟩ ‘querelle de boisson’ [‘drunken brawl’] (Grenand 1989: 222), and the Avañe’ẽ (Paraguayan Guarani) ⟨caguaí⟩ ‘atacar, atropellar, pegar’ [‘attack, run over, hit’] (Jover Peralta and Osuna 1950: 36).

Language	Form	Meaning
PTG	*- <i>kawai-p</i> /* <i>kawaiβ</i> -	‘to be aggressive, prone to violence’
Kagwahiva	<i>kag^wahiβa</i>	‘member of in-group’
Old Tupi	<i>kawaiβ</i> -	‘to be exalted, ferocious, angry’
Old Guarani	<i>kawai</i>	‘hit/beat multiple times with a club’
Avañe’ẽ	<i>kawai</i>	‘attack, hit’
Wajāpi	<i>kaʔuai</i>	‘drunk brawling’
Kamayurá	<i>kawaʔip</i>	‘member of out-group’
Kayabí	<i>kawaiβ</i>	‘member of in-group’

Table 3: Cognates of *kagwahiva*

In formal terms the comparisons are hardly problematic, which is understandable given the limited internal differentiation in the TG family. Kamayurá and Kayabí retain the final oral stop **-p* as such, with the phonetic realizations as unreleased stops [p̚] (see Weiss and Dobson 1975 and Seki 2000 for details). As mentioned before in relation to the reflexes of the PTG **kap*/**kaβ*- ‘wasp’ (see Table 2), PTG must be reconstructed with a process of lenition that changes root-final oral stops to their voiced/continuant counterparts whenever a (usually suffixal) vowel follows: thus, **-p*, **-t*, **-k* → **-β*, **-r*, **-ɣ*. This lenited alternant is that reflected as B in both Old Tupi and Kagwahiva. Finally, most dialects of Wajāpi, as well as the Guaranian varieties, have generally lost the root-final stop consonants in pre-pausal/word-final positions (see e.g. Jensen 1984, 1999).¹⁴

¹³ Note that the relevant noun phrases would be marked by the nominal function marker **-a* (see Jensen 1998, 1999). The question as to whether the relevant noun-modifier constructions are best analyzed as lexical compounds or as syntactic phrases is arguably immaterial to the proposals being made here.

¹⁴ Setting a number of complexities aside, the PTG root-final stop consonants were still retained as morphophonological alternants in Old Guarani. In Montoya’s Tesoro, for instance, the main lexical source on Old Guarani, such alternating roots are explicitly indicated. Thus, the Old Guarani reflexes of the PTG **-ʔat* ‘to fall’, **ts-endup* ‘to hear, listen’, are given, respectively, as ⟨A.r.⟩, ⟨Hendú.b⟩ (Montoya 1639: 4, 151–151v), the dot being Montoya’s convention for noting a final consonant that alternates with a ∅ preceding pause. The PTG **-k* was still not entirely lost in the language, so that reflexes of the PTG roots such as **-kutuk* ‘to pierce’ were still noted as ⟨Cutúg⟩ (Montoya 1639: 111). For Wajāpi there is, in turn, a dialectal dimension, as the

However, it is necessary to examine why a form whose core meaning is associated with violent displays of aggression might become a group's autonym. The following quotation, from a modern ethnographic description of certain aspects of Kagwahiva society, may suggest a plausible reason through the 'bellicose ethos' that characterizes the group:

The difficulties of dealing with aggressive feelings are increased by **the Kagwahiv martial tradition that makes ferocity toward the enemy as highly valued an attitude as avoidance of conflict with relatives**. One young warrior after pacification kept making good-natured but unabashed attempts to acquire the head of the first SPI agent (García de Freitas 1926: 72). Stealing, generally condemned among coresidents, was heartily engaged in during visits to Nimuendaju's pacification post (1924: 233–34). The ingenuity with which it was carried out, using specially made hooks to fish belongings from across partitions, suggested pride in such exploits. (Kracke 1978: 22–23) [emphasis mine]

Aside from such positive attitudes towards displays of violence and other examples of "uncivil" behaviour when directed against outsiders, there are other parallels that make the adoption of the reflexes of the PTG **kawaip/*kawaiβ-* as an in-group label less surprising. However, Aguilar (2017) failed to include in her review of the linguistic, historical and cultural features that bring the Kayabí closer to the Kagwahiva among the TG groups, the recognition that *kawaiwete*, the autonym of the Kayabí, is also related to the PTG **-kawaip/kawaiβ-*. Although the Kayabí employ *kawaip* as a reference to outsiders, in particular certain Rikbáktsa groups, when modified by the suffix *-ete* 'real, really, truly' (Weiss 2005: 26), the resulting *kawaiwete* (that is: *kawaip + ete*) is used as a name for in-group members. According to Stuchi (2010), *kawaiwete* was translated into Portuguese as meaning 'grande guerreiro' ('great warrior') by the Kayabí (Stuchi 2010: 27–28).

We end this section by raising one formal issue with the etymology proposed in table 3 above. The medial *-h-* in the Kagwahiva form, that is the ethnonym *kagwahiva*, implies the presence of an earlier PTG affricate, either **ts* or **tʃ*, a reconstruction not supported by any of the cognates in the remaining languages (see Carvalho 2022a on the historical phonology of the PTG affricates). I will offer a solution to this puzzling feature in the next section, while presenting a tentative etymology for the PTG etymon **-kawaip/*kawaiβ-* 'to be aggressive, prone to violence'.

5. Beyond **-kawaip/*kawaiβ-*

The etymology proposed here for the PTG **-kawaip/*kawaiβ-* 'to be aggressive, prone to violence' was first suggested by Antonio Ruiz de Montoya for this etymon's Old Guarani reflex, in the same entry of his *Tesoro* presented above. Later,

Amapari variety of the language shows a situation roughly comparable to that of Old Guarani, while the Upper Oyapock and Upper Jari varieties seem to have lost the PTG root-final stop consonants entirely.

Grenand (1989) also included a brief, yet essentially identical, proposal in her *Wajãpi* dictionary. Montoya's proposal for Old Guarani considered <Caguaí> 'to hit or beat multiple times with a club' as a compound of <Caú> 'to drink alcoholic beverages' and <ai> 'bad'. The semantic disparity between the putative derivative and the compound is enough to justify a diachronic perspective (i.e. an etymology), rather than it being a question of synchronic morphological analysis. The remainder of this section addresses a number of issues which relate, in one way or another, to the etymologization of the PTG **-kawaiip/*kawaiβ-* as stemming from the combination of **kaʔu* 'to drink alcoholic beverages' and the modifying root **-aip* 'bad/ugly; intensely'.

When reviewing the existing, published work on PTG lexical reconstruction, it is clear that any attempt at etymologization requires a fresh approach. Schleicher (1998) reconstructed **-aiβ* 'bad' on the basis of evidence from five languages only: *Wajãpi*, *Tapirapé*, *Guarayu*, *Kaiowá* and *Avañe'ẽ* (Schleicher 1998: 330). Moreover, his comparisons for the latter two languages were the (typically Guaranian) forms *βai*, which may or may not be related, but precisely because of a lack of clarity, the decision has been made not to address them at this juncture. Lemle (1971) too faced certain difficulties, namely reconstructing a doublet **aiβ*, **aiβ* with the confusing inclusion of non-cognate elements (more on this below). The latest work, that of Mello (2000), fails to clarify the situation. Two entries are provided: **aiβ* 'bad', supported by six cognates, and **aiβ* 'ugly', supported by only two cognates (Mello 2000: 151). Finally, unfortunately these studies fail to meet the necessary standards, and fail to provide appropriate sources for the comparisons, a fact that may explain some of the inaccuracies in the data (e.g. such as Mello's inclusion of the Old Tupi cognate within 'ugly' only, while the meaning 'bad' is also attested as well for the language).¹⁵

The following cognate set supports the reconstruction of the PTG etymon **-aip*, for which the meaning 'bad/ugly; intensely' is proposed:¹⁶

Language	Form	Meaning
PTG	<i>*-aip/*aiβ</i>	'bad/ugly; intensely'
Old Tupi	< <i>Aiba</i> >	'fea cousa, uil cousa' ['ugly thing, evil thing']
Tapirapé	<i>ãip</i>	'mal, ruim, feio' ['bad; spoiled; ugly']
Kamayurá	<i>a'ia'ip</i>	'intensivo' ['intensive']
Wajãpi	<i>ai</i>	'mal, mauvais' ['bad, evil']
Kayabi	<i>aip</i>	'intensificador; fraco, ruim; irreal' ['intensifier; weak; bad; unreal']

¹⁵ To be fair, Lemle (1971) should be excused from this charge, as, at the time of her writing, there were very few unpublished questionnaire forms prepared by the Museu Nacional in Rio de Janeiro that could be used to obtain the necessary data on many of the relevant languages.

¹⁶ Data sources: TUP (Drummond 1952, I 136, II, 145); TAP (Almeida, Jesus, and Paula 1983: 78); KAM (Seki 2000: 88,454); WAJ (Grenand 1989: 115); KAY (Weiss 2005: 4); GUY (Hoeller 1932: 11); OGU (Montoya 1639: 23); KAG (Betts 2012: 23); AVA (Jover Peralta and Osuna 1950: 19).

Language	Form	Meaning
Guarayu	<i>ai</i>	‘schlecht; Superlativpartikel’ [‘bad, superlative particle’]
Old Guarani	<i>ai.b</i>	‘malamente’ [‘badly’]
Kagwahiva	<i>-ahiv</i>	‘briefly with intensity’
Avañe’ẽ	<i>aiví</i>	‘ruin, vil’ [‘bad, evil’]

Table 4: Cognate set supporting the reconstruction of PTG **-aip/*-aiβ*

It is important to note that this root seems to have had, already at the PTG level, a derived and somewhat more grammatical use as an intensifier. This is suggested by the diversity of the glosses in the etymology above, including ‘evil/ugly’ (Old Tupi, Tapi-rapé, Wajãpi, Kayabí, Guarayu, Avañe’ẽ), ‘bad(ly)’ (Old Guarani), and ‘with intensity’ (Kayabí, Guarayu, Kagwahiva, Kamayurá). The presence of both meanings or uses in a single language is illustrated below with examples from Guarayu (adapted from Hoeller 1932: 11).

- (3) <che- reco ai>
1.II- life bad
‘my evil/bad life’
- (4) <i- pucu ai>
3.II- long bad
‘very long’

As to its form, a significant difference from a past proposal was noted regarding the form of this etymon. Lemle (1971: 122) reconstructed a doublet **aiβ*, **aĩβ*, with the addition of a question mark to underscore the unexplained nature of the formal variation in the second vowel (*i ~ ĩ*). The reason for Lemle’s unexplained doublet stems from her unfortunately incorrect conflation of etymologically unrelated forms: her Kamayurá comparison of, *aĩ*, one of the examples supporting the reconstruction of her *ĩ*-variant for PTG, actually belongs to a separate set, that of the PTG **ts-af̃i* ‘to hurt’.¹⁷ The reason for this confusion will be clarified below. Table 5 displays the comparative evidence for the reconstruction of the PTG **ts-af̃i* ‘to hurt’, with the cognate formatives highlighted in bold font (note that <y>, <i>, <ig>, <ĩ> and <î> area all source-specific conventions for representing [ĩ]):¹⁸

¹⁷ The Kamayurá reflex of the PTG **-aip* in Table 4 includes non-etymological glottal stops. Although the details still await a full investigation of the historical phonology, it seems Kamayurá is one of the TG languages to have resolved the hiatus in a more systematic manner by the insertion of glottal elements. Thus, the reflex of the PTG **kuwaap* ‘to know’ (Mello 2000: 174), **k^waáβ* in Schleicher (1998: 339), is *-kwahap* (Seki 2000: 460).

¹⁸ Data sources are as follows: Old Tupi (Drummond 1952, I, 106); Tenetehára (Boudin 1978: 18); Tocantins Asurini (Cabral and Rodrigues 2003: 30); Tapirapé (Praça 2007: 56); Wajãpi (Grenand 1989: 55); Kayabí (Weiss 2005: 19); Guarayu (Hoeller 1932: 38); Old Guarani (Restivo 1893: 253); Kagwahiva (Betts 2012: 24); Avañe’ẽ ‘dolor’ (Jover Peralta and Osuna 1950: 272); Mbyá (Dooley 1998: 149).

Language	Form	Meaning
PTG	* <i>ts-afɨ</i>	'to hurt'
Old Tupi	⟨ <i>Baêracîg</i> ⟩	'dor, ou dores quaisquer' ['pain, to feel pain']
Tenetehára	<i>Ahĩ</i>	'doente estar, doer, padecer' ['to be ill, to hurt, to suffer']
Tocantins Asurini	<i>-ahý</i>	'ter dor' ['to feel pain']
Tapirapé	<i>ãy</i>	'doer' ['to hurt']
Wajãpi	<i>ai</i>	'douleur' ['pain']
Kayabi	<i>-ay</i>	'doer' ['to hurt']
Guarayu	⟨ <i>Tazi, zazi, razi</i> ⟩	'Schmerz' ['pain']
Old Guarani	⟨ <i>mbaeaci, tecoaci</i> ⟩	'dolencia' ['pain']
Kagwahiva	<i>-ahy</i>	'intensely negative: hurt, pain, ache, noise'
Avañe'ẽ	<i>Tasi</i>	'dolor' ['pain']
Mbyá	<i>-axy</i>	'dor' ['pain']

Table 5: Cognate set supporting the reconstruction of PTG **ts-afɨ*

As with the reflexes of the PTG **-aip* 'bad/ugly; intensely', the reflexes of the PTG **ts-afɨ* 'to hurt' are attested throughout the family with a more grammatical use as markers for emphasis or intensification: Teneteherára *-ahy* 'enfático' (Harrison and Harrison 2013: 11); Tocantins Asurini *-a hyahy* 'intensivo' (Cabral and Rodrigues 2003: 31); Kayabí *ay* 'marcador de intensidade' (Weiss 2005: 19); Kagwahiva *-ahy* 'intensely negative: hurt, pain, ache, noise, taste' (Betts 2012: 24). This fact, however, fails to be mentioned in certain language-specific accounts. Thus, Weiss's Kayabí dictionary, while correctly featuring a single polysemic entry for *-aip* 'intensifier; weak, bad; unreal' (Weiss 2005: 4), gives two distinct entries, *-ai* 'intensity marker' and *-ai* 'to hurt' (Weiss 2005: 19) for what are, in fact, the reflexes of a single PTG form: **ts-afɨ* 'to hurt'.

With the two etymologies above acknowledged, we are now in a position to offer an explanation for the incorrect medial *-h-* noted in the preceding section when *kagwahiva* was compared to its cognates elsewhere in the family. In brief, while the glottal fricative is the standard development of the PTG **-ɸ-*, as attested in the Kagwahiva reflex ⟨*-ahy*⟩ = *-ahĩ* (< **-afɨ* 'to hurt'), the language's reflex for the PTG **-aip* shows an unexplained medial *-h-*. Given the previous (likely) use of both roots as derived intensifiers, it is unsurprising that a contamination/cross associative influence has spread the *lautgesetzlich* medial *-h-* of *-ahĩ* 'to hurt' to the Kagwahiva reflex of the PTG **-aip* 'bad/ugly; intensely', whose expected, yet unattested reflex in the language is ***-aiβ* (see Betts 2012: 23 for examples of the use of *-ahiv* as an intensifier in modern Kagwahiva). Note that a similar contamination could have operated in Avañe'ẽ, but a contamination which affects the vowel, not the consonant, of the reflex of the PTG **-aip*: ⟨*aĩvĩ*⟩ 'bad, evil' (Jover Peralta and Osuna 1950: 19) This probably reflects **-aip* when followed by a reflex of the diminutive/attenuative suffix **-i*

(see C. Jensen 1998, 1999), although it has, surprisingly, a vowel *i* reflecting the PTG **i*- which can be explained as a contamination from <Tasĩ ‘pain’ (Jover Peralta and Osuna 1950: 272), the Avañe’ẽ reflex of the PTG **ts-afɪ*.¹⁹

We have, thus, achieved a definite reconstruction of the PTG **-aip/*-aiβ-* ‘bad/ugly; intensely’, and should now address the base of the proposed *terminus a quo* for the PTG **-kawaiɪ/*kawaiβ-* ‘to be aggressive, prone to violence’, that is, **kaɫu-*, whose meaning is considered to be ‘to drink alcoholic beverages’. Initially, what was possibly the main semantic reason (that is, over and above any formal association) that may have prompted Montoya to consider the etymological association likely should be discussed. The semantic gloss offered in Montoya’s *Tesoro* for <Caguaí> (see section 4), is repeated below, but with the relevant aspect of the translation highlighted:

<Caguaí> ‘aporrear, maltratar el borracho, o el que no lo es’ [‘hit, mistreat, a drunkard, or someone who isn’t one’],

<Acaguaí hecê> ‘Maltratele’ [mistreat him/her], <Ambocaguaí> ‘hazer que le aporree’ [‘to make (someone) hit him/her’] (Montoya 1639: 85v).

The reference in Montoya to ‘a drunk person’ as the actor of the aggression is as intriguing as it is perplexing.²⁰ It is not found, for instance, in Restivo’s *Lexicon* entry for ‘aporrear’, where <Caguaí> appears. Nevertheless, it does fit exceedingly well with the cognate found in Wajãpi, at the northern limit of the TG language family and documented more than three hundred years later: *kaɫuai* ‘querelle de boisson’ [drunken brawling] (Grenand 1989: 222). Grenand herself related the form to *-kaɫu* ‘boire du cachiri’ [drink cachiri] and the root for ‘bad’ (that is, the Wajãpi reflex of the PTG **aip/*-aiβ-* ‘bad/ugly; intensely’). The base proposed for the Pre-PTG etymon underlying the PTG **-kawaiɪ/*kawaiβ-* ‘to be aggressive, prone to violence’ is a stem that can be tentatively reconstructed as **kaɫu* ‘to drink alcoholic beverages; be drunk’. Most TG languages feature a specialized verb for the meaning ‘to drink alcoholic beverages’, as in Old Guarani, where one finds: <aýú> ‘I drink water’, <acaaiú> ‘I drink mate tea’, but <acaú> ‘I drink alcoholic beverages’ (Restivo 1893: 129).²¹ The comparative evidence for the PTG **kaɫu* ‘to drink alcoholic beverages’ is given in Table 6.

¹⁹ A reviewer notes that synchronic variation between [ai] ~ [au] and [ai] ~ [aj] is common in Guarani/Avañe’ẽ, and he/she gives the example of the root <ayvu> ‘spirit; speech, speak’, for which such realizations are attested. The non-etymological *i* could, therefore, have been introduced simply by hypercorrection: *aíβi* and *aiβi* would co-exist as variants, and at some point speakers would accept that featuring *i* as the ‘basic form’. Nevertheless, the two cases are not strictly comparable: in the case of *aíβu* ‘spirit; speech, speak’ we have an etymological *i*, which, perhaps as an expression of its relatively marked status, is in fact often realized as *i* or something closer to it. I have still to see, however, any evidence of an etymological *i* shifting in realization to *i*, which is exactly what is required in the case under discussion.

²⁰ While the ordering of the constituents in Montoya’s Spanish gloss suggests that ‘the drunkard’ is the patient of the action (and the verb’s grammatical object), verbs such as *maltratar* require a preposition marking their affected arguments, unless they denote objects. We would have then *maltratar al borracho*, if the drunkard was affected by the action. This indicates that *el borracho* is intended as the agent (grammatical subject) of the action. It seems that Montoya’s gloss features *el borracho, o el que no es* not as a complement of *maltratar*, but as an appositive phrase explaining a feature of the verb’s semantics.

²¹ I am liberally translating ‘beber vino’, lit. ‘to drink wine’, as ‘to drink alcoholic beverages’.

PTG	*-kaɽu	‘to drink alcoholic beverages’
Old Tupi	⟨ka’u⟩	‘to drink alcoholic beverages, drinking spree’
Old Guarani	⟨Cáu⟩	‘to drink alcoholic beverages’
Avañe’ẽ	Ca’ú	‘to be or to become drunk; a drunkard’
Tenetehára	ka’u	‘to be drunk or intoxicated; lose control’
Guarayu	⟨ca-u⟩	‘to drink alcoholic beverages’
Kamayurá	-ka’u	‘to drink alcoholic beverages’
Wajãpi	-kaɽu	‘to drink alcoholic beverages’

Table 6: Comparative evidence for the reconstruction of the PTG *kaɽu ‘to drink alcoholic beverages’

The Old Tupi form above comes from Navarro’s Old Tupi dictionary, and represents his transliteration of the numerous textual sources from whence this form is attested (see Navarro 2013: 224). The relevant sources also record its use both as an active intransitive verb and as a noun (‘bebedeira’ [‘drinking spree’]). The hyphen ‘-’ in Guarayu ⟨ca-u, acau⟩ ‘Ich habe Chicha, Ich trinke Chicha’ [‘I have chicha, I drink chicha’]²² (Hoeller 1932: 50), and the apostrophe in the Kamayurá (Seki 2000: 238), Tenetehára (Boudin 1978: 101) and Avañe’ẽ (Jover Peralta and Osuna 1950: 41) forms stand for the same glottal stop consonant transparently indicated in the Wajãpi cognate -kaɽu (Grenand 1989: 222). Although the glottal stop usually fails to be noted in the earliest orthographies employed in the colonial period, the use of consecutive accentual diacritics in Montoya’s ⟨Cáu⟩ ‘beber vino’ (Ruíz de Montoya 1639: 95) could be interpreted as an indication of a hiatus (which was never consistently differentiated from a -V?V-sequence). The PTG *kaɽu is probably related to the reflexes of what Mello (2000: 171) reconstructs as *kawĩ ‘bebida fermentada’ [‘fermented drink’], but the precise nature of the etymological relations between these forms is unclear at the present time.²³ I mention this because, although Kagwahiva does have reflexes of *kawĩ (see Betts 2012: 129), there seems to be no attested reflex of *kaɽu in the available corpus.

The phonological adjustment required for changing *-u- into *-w- in the morpheme boundary between *-kaɽu and *-aip is not significant, as it is attested throughout the family (see e.g. ⟨u⟩ vs. ⟨gu⟩ in Guarayu ⟨acau⟩ ‘Ich trinke Chicha’ [‘I drink chicha’] (Hoeller 1932: 50), but ⟨caguar⟩ ‘Chicha-trinker’ [chicha drinker] (Hoeller 1932: 44), through the addition of the vowel-initial allomorph -ar of the agent nominalizer suffix (see also, for instance, Wajãpi *kawa* ‘buveur du cachiri’ [cachiri drinker];

²² ‘Chicha’ is one of the names used in South America for the (mostly indigenous) alcoholic beverages produced by the fermentation of tubers, such as manioc or sweet potato, or maize. Another common set of names derives from one or more reflexes of the PTG *kawĩ, as mentioned in the text.

²³ One immediate suggestion, on strictly formal grounds, would involve the widespread diminutive suffix -ĩ, but it raises semantic and morphological issues that seem more difficult to explain. Another suggestion could take *kawĩ* as the base, and derive instead *kaɽu from *kawĩ - ɽu, that is, ‘to ingest *kawĩ*’. The latter proposal faces none of the semantic and morphological problems, but it would require addressing the implications of the fact that both Kagwahiva and Kayabí lack attested reflexes of *kaɽu. I leave this for future investigations.

Grenand 1989: 223). In semantic terms, attention should be drawn to the extensions found in the Tenetehára form, documented by Boudin (1978) for the Temb  variety. The broadening from ‘(be) drunk’ to a non-specific state of absence of self-control is reminiscent of some of the changes discussed in the preceding section for the reflexes of the PTG **-kawaiip/*kawaiβ-*. With this second – and, certainly, more tentative – level of etymological complexity, the etymology proposed for the ethnonym *kagwahiva* can be summarized as follows:

Pre-PTG **-kaʔu + *-aip* ‘to drink too much; to drink badly’ > PTG **-kawaiip/*kawaiβ-*
‘to be aggressive, prone to violence’

I end by noting a phonological issue with the proposal that must still be addressed, namely the fate of the glottal stop in **-kaʔu*. Although the formation of the glide *w* in **kaʔu-aip* is predictable and lacks significance, the loss of the glottal stop is relevant, as highlighted by a reviewer. Schleicher (1998) first suggested that glottal stop clusters followed by another consonant should be reconstructed for PTG, and to achieve this one should rely specifically on the evidence from *Kagwahiva* and *Kayab *, the sole languages to retain direct evidence of these clusters. The fact that both *kagwahiva* and *kawaiip* (as well as *kawaiwete*) lack a glottal stop seems to constitute an argument against the fact that they are etymologically derived from **-kaʔu-aip*. Note, however, that preliminary evidence suggests that a number of contexts trigger the elimination of inherited glottal stops in *Kagwahiva* (see e.g. Carvalho 2022b), and this could be the case with the words under investigation. The issue as a whole requires much more discussion and argumentation than can be possibly offered within the limits of this paper, but it may have consequences for at least a subset of the claims and hypotheses advanced here.

6. Synthesis and conclusions

The immediate conclusions of the present paper have determined:

- a) The traditional etymology relating *kagwahiva* to an expression containing the noun *kaβ-* in the *Kagwahiva* language is formally untenable;
- b) The noun *kagwahiva* is actually of Proto-Tupi-Guarani provenance, a finding that adds unexpected chronological complexity to the etymology;
- c) The autonym *kawaiwete*, used by the people usually identified as *Kayab * is also a (partial) cognate of *kagwahiva*, and this is certainly a relevant fact in underscoring the relative cultural and linguistic similarities between the *Kayab * and the *Kagwahiva* (both the peoples and their languages);
- d) The medial glottal fricative *h* in *kagwahiva* results from contamination, ultimately motivated by the use of the reflexes of two independent PTG etyma, **-aip* ‘bad/ugly’ and **ts-afɿ* ‘to hurt’, as intensifiers;
- e) In semantic terms, the comparative evidence suggests an original etymon with a meaning ‘to be aggressive, prone to violence’ that was later changed to describe either the ‘uncivilized’ behaviour of out-group members or the prized warlike or aggressive disposition of in-group members towards strangers.

As advanced in section 1, the core fact grounding the proposal of a new (and, I believe, correct) etymology for the word *kagwahiva* is the existence of cognate forms elsewhere in the TG language family. That the existence of these cognates – whose relation to *kagwahiva* is at least formally straightforward – has been overlooked for so long is revealing of the developing status of the lexical and etymological studies in the Tupi-Guarani language family, even though the family is sometimes celebrated as ‘the best known genetic grouping in Amazonia’ (Jensen 1999: 125).

To an outsider, it might come as a surprise that a clearly sub-optimal etymology has been accepted as such, despite the existence of accessible comparative evidence, which, moreover, consists of easily comparable forms supported by straightforward sound correspondences. As a matter of fact, even the basic phonological reconstruction has only recently been established on a more coherent basis, both for Tupi-Guarani (Carvalho 2022a, 2022b, 2023) and for the more inclusive Tupian family (Nikulin and Carvalho 2022). There is, in fact, an urgent need for basic historical linguistic work on the Tupi-Guarani language family, and we hope that by focusing on accurate and detailed etymological studies we might be contributing to this end.

References

- Aguilar A.M. 2017. Kawahíwa como uma unidade linguística. – *Revista Brasileira de Linguística Antropológica* 9.1: 139–161.
- Almeida A., Irmãzinhas de J., Gouvêa de Paula L. 1983. *A língua Tapirapé*. Rio de Janeiro: Biblioteca Reprográfica Xerox.
- Anchieta J. de. 1595. *Arte de Grammatica da Lingoa mais usada na costa do Brasil*. [1876 facsimile print]. Leipzig: B.G. Teubner.
- Barbosa A.L. 1956. *Curso de Tupi Antigo: Gramática, Exercícios, Textos*. Rio de Janeiro: Livraria São José.
- Bačić J. 1987. Slav: The origin and meaning of the ethnonym. – *Slovene Studies* 9.1–2: 33–41.
- Betts L. 2012. *Kagwahiva dictionary*. Anápolis: Associação Internacional de Linguística SIL-Brasil.
- Boudin M.H. 1978. *Dicionário de tupi moderno: Dialeto tembê-tênêthar do alto do rio Gurupi*. [vol. 4–5] *Coleção ciências humanas*. 2 vols. São Paulo: Conselho Estadual de Artes e Ciências Humanas.
- Cabral A.S., Rodrigues Dall’Igna A. 2003. *Dicionário Asurini do Tocantins-Português*. Belém: Editora da UFPA.
- Carvalho F.O. de. 2019. On the etymology of the ethnonym ‘Katukina’. – *Revista Brasileira de Línguas Indígenas* 2.1: 5–16.
- Carvalho F.O. de. 2022a. On the Guaranian evidence for two Proto-Tupi-Guarani affricates. – *Journal of Language Relationship* 20.1: 81–112.
- Carvalho F.O. de. 2022b. A new sound change for Guarani(an): Glottal prothesis, internal classification and the explanation of synchronic irregularities. – *Folia Linguística Historica* 56: 263–288.
- Carvalho F.O. de. 2023. Proto-Tupi-Guarani did not have a palatalized velar stop. – *Boletim do Museu Paraense Emílio Goeldi* 18.1: 1–21.
- Cerrón-Palomino R. 2008. *Voces del Ande: Ensayos sobre Onomástica Andina*. Lima: Fondo Editorial de la Universidad Católica del Perú.

- Clauson G. 1963. The name Uyğur. – *Journal of the Royal Asiatic Society of Great Britain and Ireland* 3.4: 140–149.
- Dietrich W. 1990. *More evidence for an internal classification of Tupí-Guaraní languages*. [vol. 12]. Berlin: Indiana Beiheft.
- Dooley R.A. 1998. *Léxico Guaraní, Dialeto Mbyá: Versão para fins acadêmicos*. Rondonia: Sociedade Internacional de Linguística, Porto Velho.
- Drummond C. 1952. *Vocabulário na Língua Brasileira*. [2nd ed. vol. 23]. São Paulo: Etnografia Tupi-Guarani.
- Erikson P. 2004. Qu'est-ce qu'un "Ethnonyme"? L'exemple Matis (Amazonas, Brésil). – Petrich P. (ed.). *Identités: Positionnements des groupes indiens en Amérique Latine*. Paris: Le Cahiers LHIM: 127–135.
- Figueira L. 1687. *Arte de Grammatica da Lingua Brasileira*. [facsimile edition of 1880. Nova Edição. Orig. Manoel da Silva 1687 [1621]]. Rio de Janeiro: Lombaerts.
- Franchetto B., Heckenberger M. (eds.). 2001. *Os Povos do Alto Xingu. História e Cultura*. Rio de Janeiro: Editora da UFRJ.
- Grenand F. 1989. *Dictionnaire wayampi-français, lexique français-wayampi (Guyane français)*. [vol. 1.] Langues et Sociétés d'Amérique Traditionnelle. Paris: SELAF.
- Grünberg G. 2004. *Os Kaiabi do Brasil Central. História e Etnografia*. São Paulo: Instituto Socioambiental (ISA).
- Harrison C.H., Harrison C.W. 2013. *Dicionário Guajajara-Português*. Anápolis: Associação Internacional de Linguística SIL-Brasil.
- Hoeller A. 1932. *Guarayo-Deutsches Wörterbuch*. Guarayos, Bolivia. Hall in Tirol: Missionprokura der P. P. Franziskaner.
- Jensen C. 1984. O desenvolvimento histórico da língua Wayampi. [unpublished MA dissertation, Universidade Estadual de Campinas (UNICAMP)].
- Jensen C. 1998. Comparative Tupí-Guaraní morphosyntax. – Derbyshire D.C., Pullum G.K. (eds.). *Handbook of Amazonian languages*. [vol. 4]. Berlin: Mouton de Gruyter: 489–618.
- Jensen C. 1999. Tupi-Guarani. – Dixon R.M.W., Aikhenvald A. (eds.). *The Amazonian languages*. Cambridge: Cambridge University Press: 125–164.
- Jover P.A., Osuna T. 1950. *Diccionario guaraní-español y español guaraní*. Buenos Aires: Editorial Litocolor.
- Kempf B. 2011/2012. Ethnonyms and etymology – The case of Oyrat and beyond. – *Ural-Altäische Jahrbücher* 24: 189–203.
- Kracke W.H. 1978. *Force and persuasion: Leadership in an Amazonian society*. Chicago: The University of Chicago Press.
- Krivoshin de Canese N. 1983. *Gramática de la lengua Guaraní*. Asunción: Colección Ñemity.
- Lass R. 1973. Review of *The Origin of English Surnames*, by P. H. Reaney. – *Foundations of Language* 9.3: 392–402.
- Lemle M. 1971. Internal classification of the Tupi-Guarani linguistic family. – Bendor-Samuel D. (ed.). *Tupi Studies I*. [vol. 29]. Summer Institute of Linguistics Publications in Linguistics and Related Fields. Norman: Summer Institute of Linguistics of the University of Oklahoma. [available at: <http://www.sil.org/acpub/repository/11416.pdf>].
- Martius C.Ph.F. von. 1867. *Beiträge zur Ethnographie und Sprachenkunde Amerika's zumal Brasilien, II: zur Sprachenkunde*. Leipzig: Friedrich Fleischer.
- Mello A.S. 2000. *Estudo histórico da família lingüística Tupí-Guaraní: aspectos fonológicos e lexicais*. [unpublished PhD dissertation, Universidade Federal de Santa Catarina (UFSC)].
- Menéndez M.A. 1989. *Os Kawahiwa: Uma contribuição para o estudo dos Tupis Centrais*. [unpublished PhD dissertation, Universidade de São Paulo (USP)].

- Mithun M. 1991. Active/agentive case marking and its motivation. – *Language* 67.3: 510–546.
- Montoya A. Ruíz de. 1639. *Tesoro de la lengua guarani*. Madrid: Iuan Sanchez.
- Navarro E. de Almeida. 2013. *Dicionário de Tupi Antigo*. São Paulo: Global.
- Nikulin A., Carvalho F.O. de. 2022. A revised reconstruction of the Proto-Tupian vowel system. – *Boletim do Museu Paraense Emílio Goeldi* 2.17: 1–39.
- Nimuendajú C. 1924. Os Índios Parintintin do Rio Madeira. – *Journal de la Société des Américanistes* XVI: 201–278.
- Nimuendajú C. 1948. The Cawahib, Parintintin, and their Neighbours. – Steward J.H. (ed.). *Handbook of South American Indians: Vol. 3: The Tropical Forest Tribes*. Washington 1946–1950: Smithsonian Institution, Washington: Bureau of American Ethnology: 283–297.
- Pease H. 1977. Juma-Parintintin similarities. – *Arquivo Linguístico* 038: 1–7.
- Pease H., Betts L. 1971. Parintintin phonology. – Bendor-Samuel D. (ed.). *Tupi studies I*. [vol. 29]. Summer Institute of Linguistics Publications in Linguistics and Related Fields. Norman: Summer Institute of Linguistics of the University of Oklahoma. [available at: <http://www.sil.org/acpub/repository/11888.pdf>].
- Pease H. 1991. Amundava. *Formulário dos Vocabulários Padrão do Museu Nacional para Estudos Comparativos Preliminares nas Línguas Indígenas Brasileiras*. [unpublished manuscript].
- Praça W.N. 2007. *Morfossintaxe da Língua Tapirapé (Família Tupi-Guaraní)*. [unpublished PhD dissertation, Universidade de Brasília]. [available at: <http://hdl.handle.net/11858/00-001M-0000-0012-95AF-B>].
- Ramirez H. 2010. Etnônimos e Topônimos no Madeira (séculos XVI–XX): Um sem-número de equívocos. – *Revista Brasileira de Linguística Antropológica* 2.2: 179–224.
- Restivo P. 1893. *Lexicon Hispano-Guaranicum: Vocabulario de la lengua Guaraní*. Stuttgartiae: Guilielmi Kohlhammer.
- Rodrigues A.D. 1958. *Phonologie der Tupinambá-Sprache*. [unpublished PhD dissertation, Universität Hamburg].
- Sampaio W.B. de Araújo. 1998. *Estudo comparativo sincrônico entre o Parintintin (Tenharim) e o Uru-eu-uau-uau (Amondawa): contribuições para uma revisão na classificação das línguas Tupi-Kawahib*. [unpublished MA dissertation, Campinas: IEL, UNICAMP].
- Schleicher, C.O. 1998. *Comparative and internal reconstruction of the Tupi-Guarani language family*. [unpublished PhD dissertation, University of Wisconsin – Madison].
- Seki L. 2000. *Gramática do Kamaiurá: língua Tupi-Guarani do Alto Xingu*. São Paulo: Imprensa Oficial.
- Silva M.F. da. 1981. *A fonologia segmental Kamayurá*. [unpublished MA dissertation, Universidade Estadual de Campinas].
- Stuchi F.F. 2010. *A Ocupação da Terra Indígena Kayabí (MT/PA): História Indígena e Etnoarqueologia*. [unpublished MA dissertation, Universidade de São Paulo (USP)].
- Turin M. 2002. Ethnonyms and other nyms: Linguistic anthropology among the Thangmi of Nepal. – Buffetrille K., Diemberger H. (eds.). *Territory and identity in Tibet and the Himalayas*. [= *Brill's Tibetan Studies Library* 2/9]. Leiden: Brill.
- Weiss H., Dobson R. 1975. *Phonemic statement of Kayabí*. [unpublished manuscript]. [available at: <https://www.silbrazil.org/resources/archives/83772>].
- Weiss H. 2005. *Dicionário Kayabí - Português com um Glossário Português - Kayabí*. [preliminary and experimental edition]. Brasília: Summer Institute of Linguistics.