

# How to do words with ‘things’: Multiple grammaticalization from ‘thing’ in Tupi-Guarani

Johan van der Auwera, Olga Krasnoukhova

University of Antwerp, Belgium

<johan.vanderauwera@uantwerpen.be> <olga.krasnoukhova@uantwerpen.be>

In the Tupi-Guarani languages the ancestral ‘thing’ word has developed a fair number of grammatical uses, either on its own or together with other material. The paper surveys these uses and their diachronies, with respect to both general issues of grammaticalization from a ‘thing’ source or to debates specific to Tupi-Guarani languages. We first survey pronominal uses (indefinite, interrogative, and negative) and discourse particle uses. Then we turn to morphological functions serving incorporation, intransitivization and nominalization. We also deal with negative and privative functions.

KEYWORDS: grammaticalization, pronouns, indefiniteness, negation.

## 1. Introduction

This paper has three goals. First, it examines the grammaticalization that the noun ‘thing’ has undergone in Tupi-Guarani languages and it makes some new proposals. Second, it increases our understanding of the grammaticalization of ‘thing’ words in general. Third, in concert with the first two goals, it identifies some tasks for future research.

## 2. ‘Thing’ in the World Lexicon of Grammaticalization

In the second edition of the *World Lexicon of Grammaticalization* (Kuteva *et al.* 2019: 432-435), the authors list five grammaticalization pathways starting from the noun that means ‘thing’. ‘Thing’ nouns can become (i) complementizers, (ii) indefinite pronouns, (iii) attributive possession markers, (iv) nominalizers or (v) interrogative pronouns. The first three pathways were already discussed in the first edition (Heine & Kuteva 2002: 295-297). In this paper we discuss three of the five known pathways, *viz.* the ones towards indefinite pronouns, nominalizers and interrogative pronouns, and we add a few ‘new’ ones, all on the basis of the Tupi-Guarani languages.

In both editions of the *World Lexicon* no distinction is made between the grammaticalization of the ‘thing’ word occurring by itself and the case in

which the ‘thing’ word is part of a construction. Thus Swahili *kitu* ‘thing’ is said to have given rise to the indefinite pronoun ‘something’, but in Yoruba it is not *ohun* ‘thing’ by itself but the combination *ohun kan* ‘thing one’ that turned into the ‘something’ pronoun (Kuteva *et al.* 2019: 432-433). In what follows we will show that it is important to make this distinction.

The Tupi-Guarani (henceforth **TG**) branch counts 48 languages (Hammarström *et al.* 2019). Geographically, they are spoken in different parts of South America: from Guyana, Venezuela and Ecuador to Peru, Brazil, Bolivia, Paraguay, Uruguay and Argentina. TG is part of the larger Tupian family (with, according to Hammarström *et al.* 2019, 71 languages). At the end of the paper we list the TG languages that we discuss. TG forms an interesting branch for investigating the grammaticalization of ‘thing’ words. In particular, it is the incorporated and negative uses of ‘thing’ constructions that have attracted scholarly attention, especially that of Wolf Dietrich (Münster). This paper strongly relies on the work of Dietrich (1986; 1990; 1994; 2003; 2017a; 2017b). However, when it comes to the negative uses of ‘thing’, we offer partially different analyses.

In Section 3 to 5 we discuss pathways from ‘thing’ that first reach, respectively, ‘something’, ‘what’ and ‘nothing’ and then go on to other grams. Section 6 deals with the incorporated, detransitivizing, use, and Section 7 is about two nominalizing uses, both of which also take the ‘thing’ marker into the sphere of negation. Section 8 is the conclusion.

### 3. From the noun ‘thing’ to the pronoun ‘something’ and beyond

In Proto-TG the word for ‘thing’ is *\*maʔe* (Jensen 1998: 536).<sup>1</sup> This word survives in many TG languages with a similar shape and the same meaning. In some and perhaps all TG languages the word that derives from this proto-form, hence just ‘*\*maʔe* (word)’, clearly functions as a noun. Thus in Eastern Bolivian Guaraní, nominal possession is expressed with a prefix – see (1a) – and we find the same prefix on *mbáe* – see (1b).

- (1) Eastern Bolivian Guaraní (Gustafson 2014: 328; Dietrich 1994: 110)
- a. *che-reĩndĩ*  
1SG-sister  
‘my sister’
  - b. *che-mbáe*  
1SG-thing  
‘my thing’

In Teko (2) we see that the ‘thing’ word is modified by a demonstrative – just like other nouns.

- (2) Teko (Rose 2011: 237)  
*Si-ro-nan*                      *aŋ*                      *baʔe:*                      [...]
   
 1PL-CAUS.COM-run          DEM                      thing
   
 ‘We have removed this thing [...]’

In Araweté the relevant word is *meʔe*. Solano (2009: *passim*) consistently glosses the non-interrogative use as *coisa* ‘thing’, different from the interrogative use which gets a pronominal ‘what’ gloss. Nothing would indicate that the language has an indefinite pronoun equivalent to ‘something’. Similarly, in a study on interrogative words in Tupian, Brandon & Seki (1984: 94) contrast the *\*maʔe* word for ‘what’ with the non-interrogative *\*maʔe* word for ‘a thing’, and there is no mention of any pronoun ‘something’ (cf. also Cabral 1995: 211). In a later study with a similar goal the same linguists contrast the *\*maʔe* ‘what’ words with undifferentiated *algo/coisa* ‘something/thing’ words (Seki & Brandon 2007: 270 – cf. also Seki 2000: 400). Gregores & Suárez (1967: 142) have a section in their grammar of Paraguayan Guaraní with a section on ‘indefinite pronouns’, but ‘something’ is absent, while the sections on negative and interrogative pronouns do contain the *\*maʔe* forms. On the same language Dietrich (2017b: 182) is clear: Paraguayan Guaraní does not have positive indefinite pronouns like ‘something’: the work of ‘something’ in other languages is done with the noun for ‘thing’. From these statements one could infer that the relevant TG languages do not have a ‘something’ pronoun.

The meaning of ‘thing’ is, of course, very general and close to the meaning of ‘something’. In the general literature it is sometimes called a ‘generic noun’ or ‘general ontological category noun’ (both in Haspelmath 1997). In the same vein Magalhães (2007: 159) calls the noun *maʔá* in Guajá an ‘indefinite’ noun. However, for the same form *maʔe* in Tapirapé, Praça (2007: 78) explicitly says that we deal with nouns that have grammaticalized as indefinite pronouns. A similar view is implicit in Villafañe’s analysis of Yuqui *ba*: Villafañe (2004: 121) glosses its proto-form *\*maʔe* as *cosa*, *algo* ‘thing, something’, but the form *ba* is categorized as a pronoun only (Villafañe 2004: 272, 302). Neither author offers evidence, however, that the current words in Tapirapé and Yuqui are anything other than just nouns. In fact, the analysis in Praça (2007) suggests that the forms *ãwã* ‘person, people’ and *maʔe* ‘something indefinite, generic’ are ‘basically’ pronominal only in interrogative contexts:

Ainda funcionam como nomes plenos [...], mas basicamente estão sendo usados como pronomes interrogativos.  
 ‘They still work as full nouns [...], but are basically being used as interrogative pronouns.’  
 (Praça 2007: 78)

Dietrich (2003: 241) offers a similar analysis for Eastern Bolivian Guaraní: *mba?e* is translated as *cosa* ‘thing’, but in a negative context it is said to be used as the negative pronoun *nada* ‘nothing’.<sup>2</sup> Unclear also is Dooley’s (1998: 68) analysis of *mba?e* in Mbyá Guaraní: he classifies it as pronoun, but translates it as *coisa, objeto não humano* ‘thing, non-human object’, which suggests that it is a noun. For Sirionó, Priest & Priest (1985: 45) offer the opposite account: they categorize *mbae* as a noun but translate it with pronominal *algo* ‘something’.

The conclusion is that for words deriving from *\*ma?e* that do not contain any additional morphemes, no evidence has been offered so far that these are anything other than nouns. In this way the TG languages illustrate a general problem. In a 100-language sample Haspelmath (1997) has 42 languages with generic-noun-based indefinites, but, he notes, “[i]t may well be that many of these cases are not really indefinite pronouns” but just generic nouns. He further suspects that there might be subtle differences between generic-noun-based indefinites and generic nouns, but, to return to TG, we are not aware of any such differences.

For a few TG languages, a complex form is reported. Thus Mbyá Guaraní does not only have the bare *mba?e* as a potential pronoun, but also *mba?emo* (Dooley 1998: 68), in which *-mo* would be an indefiniteness marker (Dooley 1998: 21), going back to a Proto-TG word *\*amõ* ‘some’ (Lemle 1971: 116). Like the simple form, it is categorized by Dooley (1998: 68) as a pronoun, but his translation is *alguma coisa* ‘some thing’. At first sight, the translation with *alguma coisa* suggests a nominal approach, except that in Portuguese, and probably especially in Brazilian Portuguese, *alguma coisa* is grammaticalizing as a ‘multi-word’ pronoun: in informal language the phrase allows masculine agreement (Lachlan Mackenzie [Lisbon] and Nita Teixeira Da Silva [Antwerp], personal communication), instead of the feminine that the feminine noun *coisa* should trigger.<sup>3</sup>

(3) Portuguese (L. Mackenzie, N. Teixeira Da Silva, personal communication)

<i>Alguma</i>	<i>coisa</i>	<i>está</i>	<i>errad-o</i>	/	<i>errad-a</i>
some	thing	is	wrong-M		wrong-F

‘Something is wrong.’

If Dooley considers *alguma coisa* to be a pronoun, then his analysis of *mba?emo* is consistent: it would be a pronoun.<sup>4</sup> Cadogan (1992: 103-104) also considers *mba?emo* to be a pronoun (glossing it with Spanish *algo*), different from his analysis of the simple form as a noun (*cosa*).<sup>5</sup> The chances that *mba?emo* is a pronoun are indeed higher than for *mba?e*, as the difference between English *thing*, a noun, and *something*, a pronoun, suggests. Nevertheless, English *something* allows nominal uses, as in (4a), and while this is a marked use in English, in Jamaican Creole (4b) it is fully unmarked (van der Auwera & De Lisser 2019: 6).

(4) English (a) and Jamaican Creole (b)

- a. *love is this beautiful something we are all looking for in every way*  
 (< [www.bbc.co.uk/stoke/have\\_your\\_say/archive/2004/robbie/14.shtml](http://www.bbc.co.uk/stoke/have_your_say/archive/2004/robbie/14.shtml)>, accessed 29 February 2020).
- b. *an dem no andastan no iivn wan sitn niida*  
 and they NEG understand NEG even one something neither  
 ‘and they don’t understand even one thing either’

At least two other TG languages may have a complex ‘thing/something’ word containing *\*ma?e*, viz. *manungara* in Nhengatu (Da Cruz 2011: 219-220) and *ma?anuat* in Kamayurá (Seki 2007: 270). In the Nhengatu *manungara* form, which exists next to the morphologically simple *maã*, the *nungara* part must be a similarity marker ‘like’ (Dietrich 1986: 167, 317). We know that similarity morphemes appear in indefinites: we do not have to go further than English *such* with its ‘so-like’ etymology (van der Auwera & Sahoo 2020). So it is likely that the etymology of *manungara* is ‘thing-like’. *manungara* is always glossed as *alguma coisa* ‘something’ in the source (Da Cruz 2011: 47, 219, 220), which given the ambiguous nature of Portuguese *alguma coisa*, does not settle the issue whether *manungara* is a noun or a pronoun. As to *ma?anuat* in Kamayurá, we do not know the meaning of the *(?a)nuat* part. Seki (2000: 61) explicitly says that the complex form functions as a noun and the dictionary lists it as a noun (Seki 2000: 460), even though Seki (2000: 325) also categorizes it as an indefinite pronoun. So the conclusion here is, again, that there is no evidence that the complex *\*ma?e* forms are pronouns.

As to the *ma-* part of *manungara* and *ma?anuat*, it is actually not certain that it descends from *\*ma?e*. Seki (2000: 216) considers *ma-* in Kamayurá to be a word-initial indefiniteness marker (‘um formativo indefinido inicial’) and she does not suggest any link with *\*ma?e*. Neither does Rose (2011: 288) for Teko. Interestingly, Seki’s and Rose’s proposal that *ma-* is a word-initial indefiniteness marker appears in their analyses of interrogative paradigms. (5) shows the members of the interrogative paradigm of Teko as discussed and provided with etymologies by Rose (2011: 289).

(5) <i>awa</i>	‘who?’	< ‘person’/‘someone’
<i>ba?e</i>	‘what?’	< ‘thing’/‘something’
<i>manan</i>	‘how?’	< <i>ma</i> + <i>nan</i> ‘so’
<i>manam</i>	‘when?’	< <i>ma</i> + <i>na</i> ‘when’
<i>matĩ</i> <sup>6</sup>	‘where?’	< <i>ma</i> + <i>tĩ</i> LOC
<i>mananãhã</i>	‘how many?’	< <i>ma</i> + <i>nan</i> ‘so’ + <i>ãhã</i> ‘only’
<i>maŋ</i>	‘which?’	< <i>ma</i> + <i>aŋ</i> DEM
<i>ba?amõ</i>	‘why?’	< ‘thing’/‘something’ + TRANSL
<i>ba?e-r-ehe</i>	‘why?’	< ‘thing’/‘something’ + RELN + POSTP
<i>ba?e-wi</i>	‘of what?’	< ‘thing’/‘something’ + ABL

All these interrogative constructions need the question clitic. This clitic can be the last part of the constructions containing the forms listed in (5) and illustrated in example (6a), but it can also attach to another word, as in (6b).

- (6) Teko (Rose 2011: 288)
- a. *Awa=ta aŋ o-ba?e?*  
 person=Q DEM.PROX 3-do  
 ‘Who has done this?’
  - b. *Ba?e t=ede?*  
 thing Q=2SG  
 ‘What are you?’

The *ba?e* form that combines with the question clitic clearly derives from the noun meaning ‘thing’ or the pronoun meaning ‘something’. *Awa* means ‘person’ in several TG languages, it does not appear in Teko (Françoise Rose, personal communication), but one may assume that it appeared in earlier Teko. All the other forms are non-human, like *ba?e*, and they either use a form with *ba?e* or the indefiniteness marker *ma*. Given also that throughout TG *\*ma?e* often survives with the nasal bilabial *m-* instead of the plosive bilabial *b-*, it makes sense to treat the *ma-* indefiniteness marker as deriving from *\*ma?e* as well. It would then, in Teko at least, be a more grammaticalized form than *ba?e*. It is no longer a noun or pronoun and it cannot occur on its own, but it still expresses indefiniteness and it is still non-human. This also means that Teko has two formally different reflections of *\*ma?e* (and there is even a third one, viz. a nominalizer *\*ma?ē*). We come to this in Section 7.2.

Note that an allomorphy in the interrogative paradigm is found in many TG languages. In Arawete and Kamayurá we have *ma* forms in all cells except the ‘who’ cell, but in Mbyá Guaraní the ‘who’ cell uses *ma-* too. This makes sense under the hypothesis that *ma-* can lose its ontological non-human component and only express indefiniteness.<sup>8</sup>

(7)	Arawete (Solano 2009: 294-305)	Kamayurá (Seki 2000: 78, 216)	Mbyá Guaraní (Dooley 1998: 25, 65-69)
‘who?’	<i>awa</i>	<i>awa</i>	<i>mava?e</i>
‘what?’	<i>me?e / marĩ</i>	<i>ma?anuat</i>	<i>mba?e</i>
‘how?’	<i>marĩ</i>	<i>mawite</i>	<i>mba?e xa</i>
‘when?’	<i>mĩhĩjje / mĩjje</i>	<i>maramuē</i>	<i>araka?e</i>
‘where?’	<i>mamu / mĩ / mĩhĩ</i>	<i>mam / uma / umam</i>	<i>mano</i>
‘why?’	<i>marĩma / marĩmũ</i>	<i>(ma)na?are / ma?are</i>	<i>mba?e re</i>

If this analysis is correct, we touch here upon an unmistakable grammaticalization. The origin is a noun meaning ‘thing’ but the endpoint is not a pronoun meaning ‘something’, but bound morphology expressing indefiniteness.

In the above analysis, the indefiniteness marker *ma* is part of indefinite and interrogative words, but it could also occur in other words, different though related. Da Cruz (2011: 380) analyzes the Nhengatu conjunction *mairame* ‘when’ as in (8) as deriving from *maã* ‘thing’ and *rame* ‘temporal’.

(8) Nhengatu (Da Cruz 2011: 508)

*Mairame tambui ne-pratu upe ne-obrigasão*  
 when 3PL.A.CAUS.jump 2SG.S-plate LOC 2SG.S-obligation  
*re-mbau-pa(wa)*.  
 2SG.A-eat-finish  
 ‘When they put [food] on your plate, your obligation is to eat everything.’

It is hard to see how the noun ‘thing’ could itself be involved – and Da Cruz does not offer any account – but it is not hard to see that it is the ‘thing’ derived indefiniteness marker that occurs with *rame*. Nhengatu also has a hypothetical *maã* particle, analyzed by Da Cruz (2011: 502-503) as coming from *\*maʔe*, too. It can occur by itself (9a) or accompanied by the conditional marker *saa* (9b).

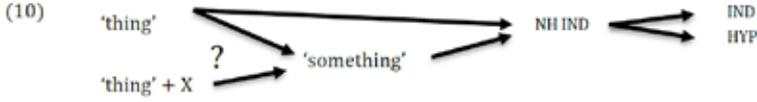
(9) Nhengatu (Da Cruz 2011: 503, 504)

a. *A-putai maã a-sendu si = nunka tau-kastigai inde*.  
 1SG.A-want HYP 1SG.A-listen COND = ever 3PL.A-punish 2SG  
 ‘I would want to hear if they ever punished you.’

b. *Saa-maá u-kwá-ramé u-pisika-maá yepesawa nhaá u-sika*  
 COND-HYP 3SG.S-know-when 3SG.A-grab-HYP first DIST 3SG.A-arrive  
*rame waá aá = pe u-yasá-rama aé*.  
 when REL DEM = LOC 3SG.S-cross-PURP 3SG  
 ‘If he knew, he would grab the first person who gets there with the intention to cross the river.’

Again, Da Cruz (2011) does not explain how a ‘thing’ noun can become a hypothetical marker. We suggest that deriving it from an indefiniteness marker makes sense.

In conclusion, it is not obvious that the TG ‘thing’ noun (either a bare noun for ‘thing’ or the noun combined with another element) has developed into a ‘something’ pronoun. It is clearer, though, that the ‘(some)thing’ (pro)noun has developed a morphological or adverbial indefiniteness use. The latter use keeps the indefiniteness component of ‘(some)thing’, but sheds either only the (pro)nominal feature (as in Teko) or both the (pro)nominal and non-human feature (as in Mbyá Guaraní). The Nhengatu hypothetical use could have developed out of the non-(pro)nominal non-human indefiniteness use.



4. From the (pro)noun ‘thing’/‘something’ to the pronoun ‘what?’ and beyond

A second grammaticalization path mentioned by Kuteva *et al.* (2019) and possibly shared by TG languages is the one that leads to the interrogative pronoun ‘what’. Kuteva *et al.* (2019) even give a TG example.

- (11) Cocama-Cocamilla (Kuteva *et al.* 2019: 435)
- |             |            |                  |                  |
|-------------|------------|------------------|------------------|
| <i>Mari</i> | <i>inu</i> | <i>yumi = ui</i> | <i>na = tsu?</i> |
| what        | 3PL.FS     | give = PST       | 2SG = DAT        |
- ‘What did they give to you?’

The example is taken from Vallejos (2016: 177). Vallejos (2016: 185) explicitly states that Cocama-Cocamilla has an interrogative pronoun derived from the noun for ‘thing’. The form is *mari*, related to the proto-form *\*ma?e* (Vallejos 2016: 614) (perhaps from Tupinambá *marā*; Cabral 1995: 215). However, there is also a noun *mari*, which appears in affirmative and negative contexts.

- (12) Cocama-Cocamilla (Vallejos 2016: 143, 100)
- (a) *Upi mari rana, yumi-n tana erutsu.*  
 all thing 3PL.MS give-NMLZ 1PL.MS bring  
 ‘All the things, which they bring, we donate.’
- (b) *Tima mari epe ey = utsu.*  
 NEG thing 2PL eat = FUT  
 ‘You will eat nothing.’

It also appears in interrogative sentences marked by the question clitic *tipa*. In most of the cases (*viz.* 5 times out of 6), however, Vallejos glosses *mari* as ‘thing’ and thus assigns the question marking exclusively to the clitic.

- (13) Cocama-Cocamilla (Vallejos 2016: 436)
- |                    |                      |
|--------------------|----------------------|
| <i>Mari = tipa</i> | <i>n = umi = ui?</i> |
| thing = Q          | 2SG = see = PST      |
- ‘What did you see?’

It is not shown, however, that the word *mari* in (13), with whatever nominal morphosyntactic properties it has, would be different from its use in (11).<sup>9</sup>

This is not to say that such morphosyntactic differences do not exist. As Françoise Rose (personal communication) points out, one would want to study possession marking, number and nominal tense. But at least one thing is clear: what makes the interrogative *mari* different from the affirmative and the negative one is its obligatory occurrence in a clause-initial position. This goes for both the use with the question particle and the one without. In their analysis of Tupian interrogatives Brandon & Seki (1984: 96-98) consider a clause-initial position to be a sufficient reason for distinguishing the interrogative from the non-interrogative *mari* type words: the two are taken as homonyms.

The demarcation problem of ‘thing’ vs ‘what?’ is by no means unique to Cocama-Cocamilla. It is a general problem for TG.<sup>10</sup> It seems that the interrogative *\*ma?e* words typically go to the clause-initial position, and so the criterion of Brandon & Seki (1984) for distinguishing between ‘thing’ and ‘what’ would be valid for all obligatorily clause-initial *\*ma?e* words.<sup>11</sup> But even if the *\*ma?e* word does not take up the clause-initial position, there is a semantic reason for taking the two to be different. The ‘what’ meaning cannot compositionally derive from the ‘thing’ meaning coupled with an interrogative marker. The compositional sense of (11) should be ‘Did they give you a thing’ or, in case Cocama-Cocamilla *mari* is a pronoun, ‘Did they give you something?’. Merely adding interrogativity to the affirmative version of (11) yields a yes/no question, and not a *wh*-question. We conclude, therefore, that in (11) but also in (13), *mari* is an interrogative pronoun, synchronically. One might object to this line of argument that it does not yet show that the *mari* of (11) and (13) is pronominal. Could it just be an interrogative ‘noun’ (cf. Brandon & Seki 1984: 97)? That would force us to an unorthodox extension of the Western canon of grammatical categories.<sup>12</sup> Traditionally, interrogative ‘what?’, ‘which?’, ‘when?’ senses are associated with ‘pro-forms’ (cf. Schachter & Shopen 2007: 33) and the ‘pro-form’ of the noun is the pronoun. So we conclude that interrogative *mari* is a pronoun.

The decision on the interrogative use of a *\*ma?e* word applies to bare *\*ma?e* uses as in (11), but no less to uses in which interrogativity avails itself of an additional marker. This may be a particle, as in Nhengatu (Da Cruz 2011: 36), or a clitic, either on the *\*ma?e* words or somewhere else in the sentence, as in Teko (6). If the clitic attaches to the *\*ma?e* word, either optionally or obligatorily, possibly as a clitic-turned-suffix, then the whole *\*ma?e* word is the interrogative pronoun.

Once the *\*ma?e* word has become an interrogative pronoun it may develop a further grammatical use. Nhengatu has a clause-final ‘protest’ particle *ba?*.

(14) Nhengatu (De Cruz 2011: 376)

*Nem*                    *maja*    *ba?*  
 CONTR.NEG    be.like    PROTEST  
 ‘There is no way.’

Da Cruz (2011: 376) suggests that it might ultimately derive from *\*maʔe*. But would it come directly from the ‘thing’ meaning? According to Da Cruz (2011: 376) this is possible, but we propose that it derives from the interrogative ‘what’. That a ‘what’ word can become a discourse marker can be seen from related and unrelated languages. In Swedish *va* ‘what’ has developed discourse marker uses and the same holds true for French *quoi* ‘what’.<sup>13</sup>

(15) Swedish (Teleman *et al.* 1998: 700; <svenska.se/tre/?sok=va&pz=2>, accessed 12 February 2020)

- a. *Du ringar mig i morgon, va.*  
 2SG ring 1SG in morning DM  
 ‘You ring me tomorrow, right?’
- b. *När jag kom hem va, fick jag se att det var  
 when 1SG came home DM got 1SG see that it was  
 översvämning i badrummet.*  
 flood in bathroom  
 ‘When I came home, right, I got to see that there was a flood in the bathroom.’

(16) French (Lefevre 2006: 110)

- Vous tombez bien. [...] Je cuisinai quelques restes.*  
 2PL fall well 1SG cooked some leftovers  
*Des palourdes en sauce, comme une fricassée quoi.*  
 of clams in sauce like a fricassee DM  
 ‘You come at the right moment. I was cooking some leftovers. Clams in a sauce, like a fricassee, right?’

In each case we have glossed the discourse marker with ‘right’, but the uses are different. In Swedish (15a) *va* adds a directive sense to what is syntactically a declarative sentence. In (15b) *va* serves the speaker to control that the hearer is paying attention. In French *quoi* – the 8<sup>th</sup> most frequent discourse marker in a list of 85 discourse markers studied by Chanet (2004) – has several functions: in (16) it is a hesitation or approximation marker. We find discourse uses in TG languages other than Nhengatu too. Priest & Priest (1985: 45) have two lemmas for *mbae*: one is a noun meaning ‘something’ and the second is the interrogative pronoun ‘what’. But there is also a lemma for a surprise particle *ba* (Priest & Priest 1985: 44). Ramirez *et al.* (2017: 482) list a ‘yes’ use under the *maʔe* ‘thing’ lemma for Warázu. To connect a ‘yes’ to ‘what’ may seem surprising, but the ‘right’ gloss used for the Swedish and French examples adds plausibility to this hypothesis. For Mbyá Guaraní *mbaʔe* Dooley (1998: 68) does not only have a lemma for the pronoun – or noun – but also one for a particle. It has two uses. In one use, it introduces a proposal with a ‘how about’ meaning, which seems easy to derive from an interrogative use. In the other use, in turn, sensibly related to the ‘how about’ use, it is paraphrased with ‘for example, perhaps’.

- (17) Mbyá Guaraní (Dooley 1998: 68)
- a. *Japytu?u mba?e.*  
rest DM  
‘How about we rest?’
  - b. *Pende-po mba?e, pende-py mba?e.*  
2PL-hand DM 2PL-foot DM  
‘Maybe it is your hand, maybe it is your foot.’

Françoise Rose (personal communication) points out that yet other candidates can be found in Guajajára (Bendor-Samuel 1972: 156), Guarayu (Hoeller 1932: 107), Kamayurá (Seki 2000: 100), Kayabí (Dobson 2005: 68) and also Guaraní Antigo (Ruiz de Montoya 1724: 188).

The two editions of the *World Lexicon of Grammaticalization* (Heine & Kuteva 2002; Kuteva *et al.* 2019) describe the development from interrogative ‘what’ to other functions, but they do not describe this path. It is not clear whether it counts as a grammaticalization, for it is not clear that a discourse particle is more ‘grammatical’ (more ‘part of grammar’) than an interrogative pronoun. Some linguists may prefer the term ‘pragmaticalization’ (cf. Degand & Evers-Vermeulen 2015) or the more general ‘constructionalization’ (Traugott & Trousdale 2013).

(18) summarizes this section. We abstract from the demarcation problem posed by ‘thing’ and ‘something’ and start the path with ‘(some)thing’.



### 5. From the (pro)noun ‘(some)thing’ to the pronoun ‘anything’/‘nothing’

It is surprising that neither Kuteva *et al.* (2019) nor Heine & Kuteva (2002) mention the negative ‘nothing’ or the negatively polar ‘anything’ sense as outcome of a grammaticalization path from ‘thing’. The textbook case is French *rien*. Its Latin ancestor was *res* ‘thing’, it means ‘nothing’ now and it has arrived to the ‘nothing’ sense through an ‘anything’ stage.

- (19) French (Mosegaard-Hansen 2013: 67)
- |                        |   |                         |   |                     |
|------------------------|---|-------------------------|---|---------------------|
| <i>Je ne dis rien.</i> | → | <i>Je ne dis rien.</i>  | → | <i>Je dis rien.</i> |
| 1SG NEG say thing      |   | 1SG NEG say anything    |   | 1SG say nothing     |
| ‘I don’t say a thing.’ |   | ‘I don’t say anything.’ |   | ‘I say nothing.’    |

We know of only one claim possibly saying that a *\*ma?e* word developed a ‘nothing’ sense. It concerns Eastern Bolivian Guaraní:

del sustantivo *mbáe* [...], empleado, en predicados negativos, como pronombre negativo ‘nada’ ‘of the noun *mbáe* [...], used, in negative predication, as the negative pronoun ‘nothing’ (Dietrich 2003: 241)

(20) exemplifies this – the example itself is from Gustafson (2012), but the glossing is based on the analysis and the comments by Dietrich (2003, personal communication).

(20) Eastern Bolivian Guaraní (Gustafson 2014: 356)

*Kuimbae mbaeti-ta mbae o-y-apo.*  
man NEG-FUT nothing 3-OBJ-do  
‘The man will do nothing.’

There are two problems with this view. First, the very fact that a ‘(some)thing’ word combines with clausal negation does not show that the ‘(some)thing’ word is itself negative. On the contrary, the combination of ‘(some)thing’ with a clausal negator makes perfect compositional sense.

(21) *It is not the case that the man will do something.*

= *The man will do nothing.*

To express the ‘nothing’ sense, the combination of a clausal negator and whatever means the language has to express ‘(some)thing’ is, in fact, the cross-linguistically most widespread strategy (van der Auwera & Van Alsenoy 2016: 483; 2018: 113).<sup>14</sup> Second, claiming that *mbae* in (20) means ‘nothing’ would mean that in a negative sentence with *mbae* negation is expressed twice, i.e. once with the clausal negator and a second time with the negative pronoun, a phenomenon known in the literature as ‘negative concord’. Dietrich (2003: 241) does not elaborate on this. Negative concord is a cross-linguistically marked strategy (van der Auwera & Van Alsenoy 2016: 483; 2018: 113), but it happens to be typical for Spanish. So in the Spanish translation we do get *nada* ‘nothing’ combined with *no* ‘not’. The better analysis of (19) has *mbae* meaning ‘(some) thing’, and this is in fact what we find in Gustafson (2012: 356): he glosses *mbae* as *cosa* ‘thing’ but in the translation in idiomatic Spanish we get *nada* ‘nothing’.<sup>15</sup>

There are at least three TG languages which are claimed to have negative concord with a *\*mba?e* word. The first two are Paraguayan Guaraní and Mbyá Guaraní, and the negative pronoun is not a bare ‘thing’ form but a complex form *mba?e-ve* (Dietrich 1994: 111; Dooley 1998: 68).<sup>16</sup>

(22) Paraguayan Guaraní (Mitāmi 2005: 60, 32)<sup>17</sup>

a. *Mitā-mi mba?eve nd-e?i-i.*  
child-DIM nothing NEG-3A.say-NEG  
‘The small child answered nothing.’

Multiple grammaticalization from ‘thing’ in Tupi-Guarani

- b. *Nda-ikatu-i kuri oi-kuaa mba?eve ambue arapy=gui.*  
 NEG-be.able-NEG PST 3A-know nothing other world = from  
 ‘He couldn’t have known anything about the other worlds.’

The negative concord seems to be strict (Estigarribia 2017: 67),<sup>18</sup> i.e. it does not depend on word order (cf. van der Auwera & Van Alsenoy 2016, 2018): in (22a) *mba?eve* precedes the negative verb, and in (22b) it follows the negative verb.<sup>19</sup> The *-ve* is hypothesized to derive from an additive meaning (‘and’, ‘also’, ‘even’) (Dmitry Gerasimov, Saint Petersburg, personal communication),<sup>20</sup> a trajectory entirely plausible on cross-linguistic grounds (Haspelmath 1997: 222-226; Van Alsenoy 2014: 228-238). This implies, as suggested in Gerasimov (2011: 71) for Paraguayan Guaraní, that the negative use must have come out of a negative polarity use, i.e. that the ancestor of *mba?eve* was felicitous in a wider set of contexts than just negative ones, such as conditionals, questions or comparatives, like English *anything*, which one can gloss as ‘even one thing’.

- (23) a. I don’t hear anything.  
 b. If you hear anything, let me know.  
 c. Do you hear anything?  
 d. I like it better than anything else.

The third language with an account implying negative concord for a *\*ma?e* word is Nhengatu, and it involves the form *nemaã*. *Nemaã* is composed of Portuguese *nem* ‘nor’ and *maã* (Da Cruz 2011: 616). Like Spanish, Portuguese has non-strict negative concord in the variant in which the clausal negator is obligatory when the negative indefinite follows the verb and forbidden when it precedes. If the examples in Da Cruz (2011) are representative, Nhengatu *menaã* negative concord is non-strict in the same way, and one suspects this to be an effect of language contact.

- (24) Nhengatu (Da Cruz 2011: 538, 187)  
 a. *Awa ti = u-puraki ti = u-riku nemaã.*  
 person NEG = 3SG.A-work NEG = 3SG.A-have nothing  
 ‘The person who doesn’t work has nothing.’  
 b. *Nemaã rupi ta-wata~wata.*  
 nothing PERL 3PL.A-RED~go  
 ‘They had nowhere to go’ (lit. ‘For nothing they went’).

(25) is a summary sketch. We draw no line between ‘(some)thing’ and ‘anything’ though it is perfectly possible, for we are not convinced that we find it in TG languages.

- (25) ‘(some)thing’  
 ‘(some)thing’ + ADD      → ‘anything’      → ‘nothing’

(25) is not quite complete. In Section 7.1 we will claim that Eastern Bolivian Guaraní *mbáeti* may have (had) a negative pronoun sense with a *\*maʔe* component. But for this we need to discuss a nominalizing use first – in Section 7.1.

## 6. From the (pro)noun ‘thing’/‘something’ to an intransitivizer

TG languages show noun incorporation (Jensen 1998: 536; Schleicher 1998: 260-267; Dietrich 1990: 304; Rose 2008) and *\*maʔe* is a noun that can be incorporated.<sup>21</sup> It is probably best described for Eastern Bolivian Guaraní, due to Dietrich (1986: 179-180; 1990: 304-307). Incorporating *mbae* into the transitive verb *a-júka* ‘kill’ yields *a-mbae-júka*, which also means ‘kill’ – or ‘hunt’ – but it cannot have a direct object (Dietrich 1990: 305). *Ja-mõi* ‘cook’ in Guayru is used when the cooking involves a specific object, but with an incorporated *mbae* it is used when the object is non-specific, in which case it cannot be expressed syntactically (Dietrich 1990: 306).<sup>22</sup> The meaning of the verb with noun incorporation can be paraphrased as ‘to thing-kill’ and ‘to thing-cook’. The incorporated ‘thing’ remains non-human; it alternates with an incorporated *\*poro* ‘person’ (Jensen 1998: 536; Dietrich 1990: 304-305), as with Eastern Bolivian Guaraní *a-poro-póta* ‘to be in love with somebody’ from *a-póta* ‘to desire’ (Dietrich 1990: 305). The incorporation is of the type that Mithun (1984) calls the ‘lexical compounding’ type.

(26)      ‘thing’       $\longrightarrow$       ‘thing’ intransitivizer

## 7. From the (pro)noun ‘thing’/‘something’ to a nominalizer and beyond

A development from ‘thing’ to a nominalizer is listed in Kuteva *et al.* (2019: 433-434). In Ewe *nú* means ‘thing’ and *nú-ɔ̀-ɔ̀* ‘thing-eat-eat’ means ‘food’ (Kuteva *et al.* 2019: 434). In TG the trajectory from ‘thing’ to a nominalizer comes in two types, both morphological, depending on whether *\*maʔe* starts or ends the derived nominal. Let us start with the *\*maʔe* element at the onset of the nominal.

### 7.1. The *\*maʔe* element at the onset of the nominal

In Dietrich (1986: 180; 1990: 310) a *mbaʔe* nominalizer is described as operating on both verbal and nominal bases: it derives both participant and situation nouns, i.e. nouns that mean either ‘thing that does/is’ or ‘situation/fact of doing/being’. On concrete nouns it has a special effect: ‘The basic word refers to determined, possessed things; the enlargement by means of *mbae-*

marks the generic, non-determined, and non-possessed character of the same notion’ (Dietrich 1994: 116). For this use Dietrich (1994: 122) contemplates considering it a generic classifier.<sup>23</sup> (27) lists some examples from Eastern Bolivian Guaraní (Dietrich 1986: 181; 1994: 114; 1990: 310; 1994: 116).

- (27) a. *kwa* ‘to know’ → *mbaékwa* ‘the one that knows’  
 b. *pōmo* ‘sticky’ → *mbaepōmo* ‘glue’  
 c. *kyra* ‘fat’ → *mbaekyra* ‘fatness, the fact of being fat’  
 d. *póty* ‘flower (of a determinate plant)’ → *mbae-póty* ‘flower (in general)’

It occurs in other TG languages too, such as Tembe (Dietrich 1994: 115):

- (28) *apu?a* ‘round’ → *ma?e-apu?a* ‘ball’

In the earlier descriptions Dietrich (1986: 180; 1990: 310) considers *mbae-* to be a prefix. In a later account, Dietrich (1994: 121) has a more nuanced view: “*mbae* is not a true nominalizing prefix [...] but [...] a lexical element whose highly unspecific meaning is most suitable for abstract, grammatical functions”. Be that as it may, we will continue to call it a ‘nominalizer’ at least, but it does seem to be less grammaticalized than the incorporated *mbae-* (Dietrich 1994: 113). Dietrich (1994: 113) points out that the incorporated *mbae* has a nasal allomorph *ma?e* in a nasal context, the noun-initial nominalizer does not, and he interprets the nasality as sign of a loss of autonomy. On the other hand, the noun-initial nominalizer lost the restriction to non-humans (Dietrich 1990: 313), as illustrated in (27a); the incorporated *mbae* did not.

A special use, we claim, is the one that leads to the Eastern Bolivian standard negator *mbáeti*.<sup>24</sup>

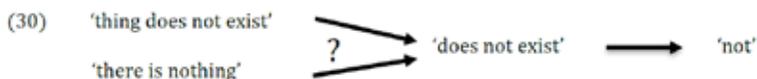
- (29) Eastern Bolivian Guaraní (Dietrich 2003: 241)

<i>Mbáety</i>	<i>o-kýje</i>	<i>chú-gui.</i>
NEG	3-get.scared	3-SEP
‘He was not scared of him.’		

*Mbáeti* is composed of a form of *mbae* and something else. In Dietrich’s latest analysis (2017a: 15) the *mbae* component is the noun ‘thing’ and it combines with *n-i-týb-i* ‘NEG-3-amount-NEG’. The resultant meaning would be ‘thing does not exist’ or ‘there is nothing’. This, in turn, would have developed into an existential negator and from there to a standard negator.

This analysis is problematic, however. It is correct that existential negators can develop into standard negators (Croft 1991; Veselinova 2014; Hamari & Veselinova eds. 2021), i.e. that ‘does not exist’ turns into ‘not’. However, the pathway sketched by Dietrich (2017a) is different: it starts from ‘thing

does not exist’ or ‘there is nothing’. How does one get from the latter to ‘does not exist’?



Second, assuming *mbae n-i-týb-i* meant ‘thing does not exist’ or ‘there is nothing’ and that *mbae* is ‘thing’, it follows that *n-i-týb-i* must be a negative existential by itself. This view is also supported by other TG languages. Here is the list of existential negators provided by Dietrich (2017a: 14-16). There is not a single one that contains a clear *mbae* component.

(31) Xingú Asuriní	<i>natyvi</i>	Kamayurá	<i>anite</i>
Old Omagua	<i>nati</i>	Guajá	<i>na'axi</i>
Nhengatu	<i>niti</i>	Teko	<i>dati</i>

Third, if we now look in more detail at the *n-i-týb-i* part, we have to ask how it could come to express existential negation. The *n-...-i* part is not a problem. It must derive from a bipartite Proto-TG negator *\*ni-...-i* (Jensen 1998: 545). The problematic part is therefore the middle part. In his latest analysis, Dietrich (2017a: 15) derives it from ‘3-amount’. The path from ‘3-amount’ to ‘exist’ is certainly not obvious. Neither is the path suggested by Dietrich’s earlier hypothesis (1986: 144; 2003: 241), in which the middle part *týb* derives from a verb *hetyp* ‘not want, not agree’.<sup>25</sup> The easiest hypothesis would be to assume that the middle part basically means ‘exist’. Teko may provide support for this view. Dietrich (2017a: 14) implies that the Teko negative existential *dati* has the same origin as the one posited for Eastern Bolivian Guaraní, but there is an easier solution. In *dati* we see the bipartite negator *d-...-i* (Rose 2011: 190), from the proto-form *\*n(a/i)-...-i* (Jensen 1998: 545), Teko has ‘be’ verb *tui* (Rose 2011: 185), and we tentatively suggest that *týb* derives a verb that relates to the modern Teko *tui* verb.<sup>26</sup> But this is not the end of the story.

In the preceding paragraph we were sceptical about the path from ‘3-amount’ to ‘exist’, but that does not imply that the ‘3-amount’ analysis itself is not possible. It could be right and, in combination with *n-...-i*, the whole thing would mean something like ‘does not exist in quantity’ – a paraphrase suggested by Dmitry Gerasimov (personal communication). At this stage we do not choose between the ‘does not exist’ and ‘does not exist in quantity’ analyses.

We still have to explain how ‘does not exist’ or ‘does not exist in quantity’ combines with a form of *mbae* ‘thing’ to ultimately yield the standard negator. Dietrich cannot help us here. In the latest analysis Dietrich (2017a: 15) suggests that the *mbae* bit derives from the noun *mbae* ‘thing’. In the ‘middle’ analysis

(Dietrich 2013: 251) the source is again identified as the noun but he adds that this noun is used as a negative indefinite ‘nothing’ in negative context. In Section 5 we have cast doubt on this view. In the oldest analysis (Dietrich 1986: 144), the analysis is different, but a bit vague. The *mbae* noun, which is the probable origin of the *mbae* bit of *mbaeti*, relates either to the noun-initial nominalizing use that we just discussed or to a privative noun-final use. We will discuss the privative use in the Section 7.2, but since this *mbae* is word-final, it is unlikely to have much to do with the word-initial use that we find here. The first hypothesis, however, we argue, is the right one. If ‘thing which is round’ gives us ‘ball’, as in (28), then ‘thing which does not exist’ or ‘thing which does not exist in quantity’ would give us ‘nothing’ or ‘little’.

Then we have to explain how ‘nothing’ or ‘little’ can become ‘not’. We start with ‘nothing’ and we do not have to go further than English. In English the change took place in sentences which did not only contain the ‘nothing’ word (the ancestor of *not*) but a standard negator as well (*ne* in older English). What happened then was that the standard negator disappeared and the ‘nothing’ word became the new negator.

(32) *ne* → *ne ... not* → *not*  
 NEG            NEG ... nothing            NEG

In Latin a negative word like ‘nothing’ changed to a standard negator without the presence of an older negator: the standard negator *ne* merged with *oenum* ‘one’, becoming ‘none’, and the resulting *noenum* became the standard negator *non*.<sup>27</sup>

(33) *ne* → *ne ... oenum* → *non*  
 NEG            NEG ... one            NEG

In both cases, i.e. the English and the Latin case, we can consider the development to be types of a ‘Jespersen Cycle’ (van der Auwera, Krasnoukhova & Vossen, 2021) and in both cases ‘nothing’ goes to ‘not’ via an intermediate emphatic stage of the marker meaning ‘not at all, in no way’. It is hard to say what road Eastern Bolivian Guaraní could have taken. An example such as (34) suggests that the language followed the same road as English. Dietrich glosses *mbáeti* as NEG but paraphrases it with *nada / en absoluto* ‘nothing / absolutely’.

(34) Eastern Bolivian Guaraní (Dietrich 2003: 241).  
*Ai-kuā-a mbáety.*  
 1SG-know-NEG NEG  
 ‘No lo conozco nada / en absoluto.’  
 ‘I don’t know him at all.’



(37) Yuqui (Villafañe 2004: 178, 183)

- a. *Baraguasu-biti ño-che a-cha-ta.*  
 rice-NEG FOC-1SG 1SG-buy-FUT  
 ‘It is not rice that I will buy.’
- b. *Ma kiabuso a-u-ta jiri.*  
 NEG banana 1SG-eat-fut NEG  
 ‘I will not eat banana.’

We have no proposal as to how *biti*, assuming that it is indeed related to *mbáeti*, is to be put on the path sketched in (36). Whereas Eastern Bolivian Guaraní and Tapiete are closely related, Yuqui is genetically further removed. Dietrich (2003: 245) suggests that *biti* could be due to ancient language contact with Eastern Bolivian Guaraní.

### 7.2. The \*maʔe element at the end of the nominal

In the preceding section the \**maʔe* nominalizer occurred at the beginning of a TG word, but it can also occur at the end of the word. This is illustrated in (38) and (39).

(38) Guajá (Magalhães 2007: 64)

- Awá-wanihã wata-maʔá.*  
 Guajá-man go-NMLZ  
 ‘The Guajá man is a walker/hunter.’

(39) Tapirapé (Praça 2007: 71)

- Akomaʔe i-eew-amaʔe n = a-ãpa-j ka.*  
 man 3-be.lazy-NMLZ NEG = 3-make-NEG farm  
 ‘The man who is lazy does not farm.’

The noun-final nominalization is productive throughout the family (Schleicher 1998: 136ff; Jensen 1998: 542-544), but there are restrictions. First, it seems to act only on verbal bases – and not nominal ones – and it may be seen, at least in some languages, as acting on an entire clause rather than just its verb or predicate (Jensen 1998: 543; Rose 2011: 145-146). Second, it seems to function mostly for participant nominalization, to the extent that Rose (2011: 145-146, 367) considers the notion of ‘nominalization’, proposed by Jensen (1998: 542-543), to be too wide and that it had better be replaced by ‘relativization’, at least for Teko (cf. Rodrigues 1953: 147-148). W. Dietrich (personal communication) is of the same opinion. However, one then has to accept that the \**maʔe* marker in Teko, viz. *-maʔẽ*, does not only function as a relativizer, but also as a complementizer.

(40) Teko (Rose 2011: 351)

*A-potane am pita-kom lekol-a-pe o-ike-mā?ē.*  
 1SG-want here child-PL school-REF-LOC 3-enter-COMPL  
 ‘I wish that the children enter the school only here.’

A complementation analysis makes as much sense as one in terms of a situation nominalization, viz. one of children entering the school. For this reason, yet without wanting to settle the dispute (cf. also Van Gijn *et al.* 2015), we will maintain Jensen’s wide concept of nominalization.

Languages may have both the noun-initial and noun-final types. (41) is a tentative sketch of what we find in a few languages.

(41) Noun-initial nominalization	Eastern Bolivian Guaraní, Guarayo, Sirionó.	Dietrich 1986: 180-181; 1990: 304-306; 1994: 114-119.
Noun-final nominalization	Kayabí, Kamayurá, Tapirapé, Guajajára, Tupinambá, Teko.	Dietrich 1994: 111; Jensen 1998: 542-543; Rose 2011: 367-368.
Noun-initial and noun-final nominalization	Mbyá, Tembe, Urubú-Kaapor, Wayampi.	Dooley 1998: 67, 112; Dietrich 1994: 111-119; Jensen 1998: 542-543.

This needs to be investigated further, as is the exact shape of the two nominalizers. In Mbyá, for instance, they are different: Dooley (1998: 112) gives *ma?e/mba?e* for the noun-initial form but *va?e* for the noun-final one.

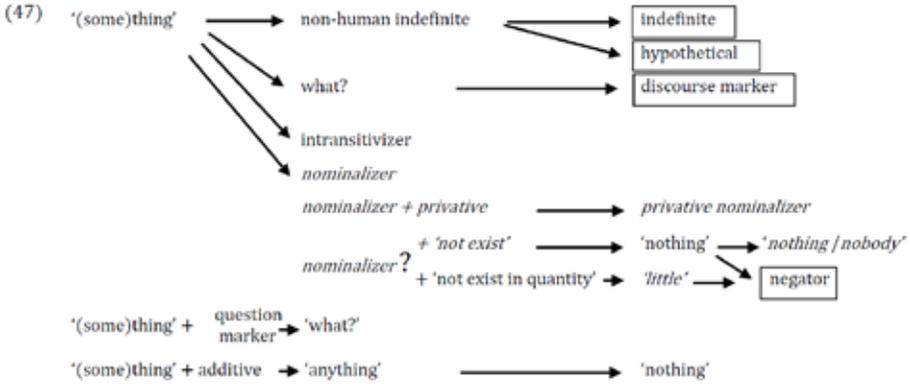
The strange thing is that in a few TG languages the *\*ma?e* word-final nominalizer has a negative sense, more particularly, a privative one. Dietrich (2017a: 18) lists Yuki, Sirionó and Eastern Bolivian Guaraní.

(42) Eastern Bolivian Guaraní (Dietrich 1986: 178)

*mbia i-yiwa-mbáe*  
 man 3-arm-nothing  
 ‘a man without arms’

How can this nominalizer be negative in this language? In his early work (1986: 178; 1990: 303) Dietrich analyzes this *mbáe* as a negative pronoun ‘nothing’, but we argued in Section 4 that this does not exist. Dietrich (2003: 242-243) still mentions this view but he considers another hypothesis (due to personal communication with Aryon Dall’Igna Rodrigues) to be “very convincing”, viz. the idea that the privative sense, as in (42), developed from a merger of the Proto-TG privative *\*e?ým*<sup>30</sup> and the nominalizer *\*ma?e*. We see a combination of these elements – in the ‘right’ order – in (43) and (44), assuming for (43) that *-(w)a?e* derives from *\*mba?e*.





From the proposed diachronic developments in (47) at least three have been discussed in grammaticalization studies, i.e. a development of ‘thing’ into an indefinite pronoun, an interrogative pronoun, or a nominalizer. As far as we know, the other developments outlined in (46) have not received attention so far. Note that (47) leaves it vague whether TG ‘thing’ has turned into an indefinite pronoun corresponding to English ‘something’. At least up to now TG scholarship does not show clear evidence for this. This is a task for future work. The main future task is to try to find direct diachronic evidence for the changes we have suggested, here based on synchronic comparison and on the typological literature. A special issue is that of the origin of the middle part of *mbáeti*, a point for which we allowed two hypotheses, i.e. a ‘nothing’ and a ‘little’ hypothesis, with a preference for the first one. Prosentential negatives require more attention too. Some other matters need additional synchronic research. We need a better analysis of negative concord (strict or non-strict and, if non-strict, what subtype?). The discourse uses also need more work – they are mentioned for a few languages, but chances are that they have stayed under the radar for others. Finally, there is no claim that the sketch is complete, in the sense that there are no further significantly different uses. Thus Dmitry Gerasimov (personal communication) mentions two further uses in Paraguayan Guaraní: (i) *ha-mbaʔe*, literally ‘and-thing’, which like English ‘and things/stuff’ is used as a ‘general extender’ (Mauri 2017: 315-318), and (ii) *maʔe-rã* ‘whatchamacallit’ (‘what you may call it’), in which *-rã* is a nominal future suffix. In *ha-mbaʔe* the *-mbaʔe* is no doubt a noun (and not a pronoun), but in *maʔe-rã* the *maʔe* part may well be the interrogative pronoun, given what we know about ‘whatchamacallit’ constructions cross-linguistically (Haspelmath 1997: 130-133).

*TG languages discussed in this paper*

We use the language names, orthographies and codes of <glottolog.org>

(Hammarström *et al.* 2019). If the name of the language listed in a source is unrecognizably different from the *Glottolog* name, we list that as well.

Araweté, araw1273	Sirionó, siri1273
Cocama-Cocamilla, coca1259	Suruí do Pará, suru1261
Eastern Bolivian Guaraní, east2555,	Tapirapé, tapi1254
Chiriguano	Teko, emir1243, Emerillon
Guajá, guaj1256	Tembe, temb1279
Guajajara, guaj1255	Tupinambá, tupi1273, língua geral
Guarayu, guar1292	amazônica
Kamayurá, kama1373	Urubú-Kapor, urub1250
Kayabí, kaya1329	Wayampi, waya1270
Mbyá Guaraní, mbyy1239	Warázu. paus1255
Nhengatu, nhen1239	Xingú Asuriní, xing1248
Old Omagua, omag1248	Yuqui, yuqu1240
Paraguayan Guaraní, para1311	

#### *Abbreviations*

1 = first person	LOC = locative
2 = second person	M = masculine
3 = third person	MS = male speech
A = active	NEG = negation
ABL = ablative	NH = non-human
ACC = accusative	NMLZ = nominalizer
ADD = additive	OBJ = object
CAUS = causative	PERL = perlative
COND = conditional	PL = plural
COM = comitative	POSTP = postposition
COMPL = completive	PRIV = privative
CONTR = contrastive	PROTEST = protest marker
DAT = dative	PROX = proximal
DEF = definite	PST = past
DEM = demonstrative	PURP = purpose
DIM = diminutive	Q = question
DIST = distal	RED = reduplication
DM = discourse marker	REF = referential
EPEN = epenthetic	REL = relative
F = feminine	RELN = relational
FS = female speech	SEP = separative
FUT = future	S = stative
HYP = hypothetical	SG = singular
IND = indefinite	TRANSL = translative
INTR = intransitivizer	

## Acknowledgements

This work was financed by the Research Foundation Flanders and is part of a larger research effort on negation in the indigenous languages of South America. Our gratitude goes to a few people, mentioned in the text, but especially to Pier Marco Bertinetto (Pisa), Wolf Dietrich (Münster), Dmitry Gerasimov (Saint Petersburg), Françoise Rose (Lyon) and Maura Velázquez-Castillo (Fort Collins, Colorado) for their most helpful and detailed comments.

## Notes

<sup>1</sup> Jensen (1998) represents the glottal stop with an apostrophe. Here and elsewhere we represent it with 'ʔ'. We represent the proto-form with a nasal, but perhaps there was allomorphy already in the proto-language (Jensen 1998: 542; Rose 2013: 43).

<sup>2</sup> We will come back to the use of the 'thing' words in interrogative and negative contexts in Sections 4 and 5.

<sup>3</sup> As Pier Marco Bertinetto points out, Italian *qualche cosa* 'some thing' followed this trajectory too, and it went further. It led to the form *qualcosa* 'something', which elided the *-che-* part. This way it surpasses English *something*.

(a) Qualche cosa è sbagliata nel mio ragionamento.  
some thing.F is mistaken.F in.DEF my reasoning

(b) Qualcosa è sbagliato nel mio ragionamento.  
something is mistaken.M in.DEF my reasoning  
'Something is wrong in my reasoning.'

<sup>4</sup> In his latest account (Dooley 2006: 107) only lists the complex *mba?emo* as an indefinite pronoun, not the simple *mba?e*.

<sup>5</sup> Martins (2003) only mentions the simple form and treats it as a noun.

<sup>6</sup> The *matì* part also occurs in the word *matirane* 'somewhere, anywhere' (Rose 2011: 266).

<sup>7</sup> Rose (2011: 288) gives 'Who are you?' as a translation.

<sup>8</sup> In (7) we do not mark whether the construction needs a question particle (like in Teko), forbids it, like in Kamayurá, or allows it (like in Tembe). A related parameter is whether or not the forms can occur by themselves in non-interrogative contexts. See Brandon & Seki (1984) and Seki & Brandon (2007) on these parameters of variation.

<sup>9</sup> It is also not discussed whether *mari* spawned an indefinite 'something' pronoun.

<sup>10</sup> It also shows up in the general literature on indefinite and interrogative pronouns, cf. Bhat's (2004: 225) 'Interrogative-Indefinite puzzle'.

<sup>11</sup> One could imagine a TG language in which both the interrogative and the non-interrogative words are slightly different, though both derive from *\*ma?e* and are not part of a more complex word. We do not know of any such TG language.

<sup>12</sup> Of course, there is nothing wrong with adding unorthodox categories. Relative to the indefinite-interrogative demarcation problem, there have already been some proposals, viz. 'ignorative pronoun' (Karcevski 1969: 216) or 'episteme' (Mushin 1995), categories that are to capture what interrogatives and indefinites have in common. As argued above, the semantic non-compositionality of the interrogative uses of *\*ma?e* words plead against these notions.

<sup>13</sup> Special thanks are due to Hartmut Haberland (Roskilde) and Patrick Dendale (Antwerp) – and his web page <www.uantwerpen.be/en/projects/lexicales/recherche-dans-la-bibliographie>, accessed 2 March 2020 – for helping with the relevant references.

<sup>14</sup> *The man will not do something* also has a pragmatically marked use, with 'something' scoping over negation, resulting in a 'there is something that the man will not do' reading, but the fact that the most widespread expression strategy for 'nothing' combines '(some)thing' with the clausal negator shows that this marked reading has little impact on the grammar.

<sup>15</sup> Wolf Dietrich (personal communication) points out that the arguments do not rule out that a *\*mba?e* word that cooccurs with clausal negation has the negatively polar 'anything' sense. We agree and we know that this could be argued for Spanish *nada* as well. However, for *\*mba?e* we lack positive evidence, different from what we have for Paraguayan Guaraní *\*mba?eve* (see further on).

<sup>16</sup> We do not deal with the larger question whether TG languages have negative concord with other words, like *peteĩ* 'one' in *nipeteĩ* 'not one (thing/person)' (Gregores & Suárez 1967: 142).

<sup>17</sup> Thanks are due to Bruno Estigarribia (Chapel Hill) and Wolf Dietrich (Münster) for help with these examples.

<sup>18</sup> Estigarribia (2017: 67) mentions that there is no concord with the phasal adverb *ne?ĩra* 'yet', in the sense that *ne?ĩra* is different from negative pronouns and does not cooccur with the verbal negator. Gerasimov (2011: 70) may explain why this is the case: *ne?ĩra* is itself negative ('not yet') and with respect to negative concord it functions like a verbal negator (thus also triggering concordant negative pronouns).

<sup>19</sup> In Spanish negative concord is non-strict: preverbal negative indefinites forbid negative concord but postverbal ones require it.

(a) *Nadie* (\*no) *admira* a *Carlos*. (b) *Carlos* \*(no) *admira* a *nadie*.  
 Nobody NEG admires to Charles Charles NEG admires to nobody  
 'Nobody admires Charles.' 'Charles admires nobody.'

Kallfell (2011: 120; 2016: 15) points out, with reference to Krivoshein de Canese & Corvalán (1987: 72-73), that Standard Paraguayan Spanish relaxes the ban on negative concord with preverbal indefinites, under the influence of Paraguayan Guaraní.

<sup>20</sup> Earlier, Gerasimov (2011: 70) related *-ve* to a universal quantification. This is not at odds with the current proposal, for universal quantification and addition/conjunction are themselves related (see e.g. Gil 1993).

<sup>21</sup> Brandon & Seki (1984: 94) also mention a 'thing' word other than *\*mba?e*, viz. *kat*, incorporated in a few TG languages in their intransitive 'eat' verb, composed of *kat* and a *-?u* root meaning 'eat'.

<sup>22</sup> In Teko the non-specific non-human object is expressed with a morpho-syntactically 'normal' 'thing' noun (Rose 2011: 176).

<sup>23</sup> The use of a 'thing' marker seems to be a strategy not distinguished yet in typological studies on genericity (Behrens 2000, 2005). Thanks are due to Leila Behrens (Köln) and Jan Rijkhoff (Aarhus).

<sup>24</sup> This is the spelling in Dietrich (1986: 144). Other spellings are *mba?eti*, *mbaéty* and *mbáety*. In our examples we will use the source spellings. In the text we use *mbáeti*.

<sup>25</sup> There is a difference between the 1986 and the 2003 analysis. In the 1986 analysis the 'not want, not agree' conjecture is part of a negative existential analysis; this is not the case in the 2003 analysis.

<sup>26</sup> For Old Omagua *nati* O'Hagan (2011: 104) suggests a distal demonstrative for the middle part.

<sup>27</sup> Or consider the standard negator *llëme* in the Chibchan language Teribe, which

arguably derives from *llë* ‘thing’ and *-me* ‘not’ (van der Auwera & Krasnoukhova 2020: 9).

<sup>28</sup> González (2005: 251) glosses *mba?eti* in (35a) with ‘anything’, which is surprising in two ways: (i) (34a) has no clausal negator; so the negative meaning must be due to the pronoun, and (ii) ‘thing’ is non-human but the sense is human.

<sup>29</sup> We owe this reference to Wolf Dietrich (personal communication).

<sup>30</sup> Depending on the language *\*e?ým* also has non-privative uses (Schleicher 1998: 274-299; Jensen 1998: 547).

<sup>31</sup> The gloss is actually NEG, but is clear that the negation that is meant is the privative one.

### *Bibliographical References*

- Behrens, Leila 2000. *Typological parameters of genericity*. Köln: Institut für Sprachwissenschaft, Universität zu Köln.
- Behrens, Leila 2005. Genericity from a cross-linguistic perspective. *Linguistics* 43. 275-344.
- Bendor-Samuel, David 1972. *Hierarchical structures in Guajajara*. Norman: The Summer Institute of Linguistics and the University of Texas at Arlington.
- Bhat, Darbhe Narayana Shankara 2004. *Pronouns*. Oxford: Oxford University Press.
- Brandon, Frank Roberts & Seki, Lucy Ferreira 1984. Moving interrogatives without an initial +WH node in Tupí. In Cook, Eung-Do & Gerds, Donna B. (eds.), *Syntax and semantics. Vol. 16. The syntax of Native American Languages*. Orlando: Academic Press. 77-103.
- Cabral, Ana Suelly De Arruda Camara 1995. *Contact-Induced Language Change in the Western Amazon: the Non-Genetic Origin of the Kokama Language*. PhD dissertation, University of Pittsburgh.
- Cadogan, Leon 1992. *Diccionario mbyá-guaraní-castellano*. Asunción: Centro de Estudios Antropológicos de la Universidad Católica.
- Chanet, Catherine 2004. Fréquence des marqueurs discursifs en français parlé: quelques problèmes de méthodologie. *Recherches sur le français parlé* 18. 83-106.
- Cocaud-Degrève, Oscar 2019. Negation in Bolivian Guaraní. Paper presented at the Atelier Morphosyntaxe, Lyon 2, December 2019.
- Croft, William 1991. The evolution of negation. *Journal of Linguistics* 27. 1-27.
- da Cruz, Aline 2011. *Fonologia e Gramática do Nheengatú: A língua geral falada pelos povos Baré, Warekena e Baniwa*. PhD dissertation, Vrije Universiteit Amsterdam.
- Degand, Liesbeth & Evers-Vermeul, Jacqueline 2015. Grammaticalization or pragmaticalization of discourse markers? More than a terminological issue. *Journal of Historical Pragmatics* 16. 39-85.
- Dietrich, Wolf 1986. *El Idioma Chiriguano: Gramática, textos, vocabulario*. Madrid: Instituto de Cooperación Iberoamericana.
- Dietrich, Wolf 1990. Chiriguano and Guayano. In Payne, Doris L. (ed.), *Amazonian linguistics. Studies in Lowland South American languages*. Austin:

- University of Texas Press. 293-320.
- Dietrich, Wolf 1994. Word formation, syntax, or noun classification? Tupi-Guarani *mba'ë* – between lexicon and grammar. *Revista latinoamericana de estudios etnolingüísticos* 8. 109-124.
- Dietrich, Wolf 2003. Formas de la negación en las lenguas tupí-guaraníes de Bolivia. *Thule. Rivista italiana di studi americanistici* 14/15. 233-250.
- Dietrich, Wolf 2017a. Tipologia morfossintática da negação nas línguas do tronco Tupi. *Liames* 17. 7-38.
- Dietrich, Wolf 2017b. Word classes and word class switches in Guaraní syntax. In Estigarribia, Bruno & Pinto, Justin (eds.), *Guarani linguistics in the 21st century*. Leiden: Brill. 158-193.
- Dobson, Rose 2005. *Aspectos da Língua Kayabi*. Cuiabá: Sociedade Internacional de Lingüística.
- Dooley, Robert A. 1998. *Léxico Guarani, Dialeto Mbyá: Versão para fins acadêmicos*. Porto Velho: Sociedade Internacional de Lingüística.
- Dooley, Robert A. 2006. *Léxico Guarani, Dialeto Mbyá com informações úteis para o ensino médio, a aprendizagem e a pesquisa lingüística. Introdução: informações gerais, esboço gramatical e referências*. <www.sil.org/resources/archives/16958>, accessed 26 March 2020.
- Estigarribia, Bruno 2017. A grammar sketch of Paraguayan Guaraní. In Estigarribia, Bruno & Pinto, Justin (eds.), *Guarani linguistics in the 21st century*. Leiden: Brill. 7-85.
- Gerasimov, Dmitry V. 2011. Otrizatelnyje mestoimenija v Paraguayskom Guarani [Indefinite pronouns in Paraguayan Guaraní]. *Acta Linguistica Petropolitana* 7,3. 67-73.
- Gil, David 1993. Conjunctive operators: areal phenomenon or semantic universal. In Gil, David (ed.), *Studies in number and quantification*. Konstanz: University of Konstanz.
- González, Hebe A. 2005. *A grammar of Tapiete (Tupi-Guarani)*. PhD dissertation, University of Pittsburgh.
- Gregores, Emma & Suárez, Jorge A. 1967. *A Description of Colloquial Guaraní*. Berlin: Mouton de Gruyter.
- Gustafson, Bret 2014. Guaraní. In Muysken, Pieter & Crevels, Mily (eds.), *Lenguas de Bolivia, III Oriente*. La Paz: Plurales. 307-368.
- Hamari, Arja & Veselinova, Ljuba (eds.) 2021. *The Negative Existential Cycle from a historical-comparative perspective*. Berlin: Language Science Press.
- Hammarström, Harald; Forkel, Robert & Haspelmath, Martin 2019. *Glottolog 4.1*. Jena: Max Planck Institute for the Science of Human History. <glottolog.org>, accessed 26 March 2020.
- Haspelmath, Martin 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Heine, Bernd & Kuteva, Tania 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Hoeller, Alfredo 1932. *Guarayo-Deutsches Wörterbuch, Guarayos, Bolivia*. Hall in Tirol: Missionprokura der P. P. Franziskaner.
- Jensen, Cheryl 1998. Comparative Tupí-Guaraní Morphosyntax. In Derbyshire, Desmond C. & Pullum, Geoffrey K. (eds.), *Handbook of Amazonian Languages, Vol. IV*. Berlin: Mouton de Gruyter. 487-618.
- Kallfell, Guido 2011. *Grammatik des Jopara. Gesprochenes Guaraní und Spanisch in*

- Paraguay. Frankfurt am Main: Peter Lang.
- Kallfell, Guido 2016. *¿Cómo hablan los paraguayos con dos lenguas? Gramática del jopara*. Asunción: Centro de Estudios Antropológicos de la Universidad Católica 'Nuestra Señora de la Asunción'.
- Karcevski, Serge 1969. Sur la parataxe et la syntaxe en russe. In Godel, Robert (ed.), *A Geneva School Reader*. Bloomington: Indiana University Press. 212-217.
- Krivoshein de Canese, Natalia & Corvalán, Graziella 1987. *El español del Paraguay. En contacto con el guaraní*. Asunción: Centro Paraguayo de Estudios Sociológicos.
- Kuteva, Tania; Heine, Bernd; Hong, Bo; Long, Haiping; Narrog, Heiko & Rhee, Seongha 2019. *World Lexicon of Grammaticalization*. 2nd edition. Cambridge: Cambridge University Press.
- Lefevvre, Forence 2006. *Quoi de neuf sur quoi? Étude morphosyntaxique du mot quoi*. Rennes: Presses Universitaires de Rennes.
- Lemle, Miriam 1971. Internal classification of the Tupi-Guarani linguistic family. In Bendor-Samuel, David (ed.), *Tupi studies I*. Norman: Summer Institute of Linguistics of the University of Oklahoma. 107-129.
- Magalhães, Marina Maria Silva 2007. *Sobre a morfologia e a sintaxe da língua Guajá (Família Tupí-Guaraní)*. PhD dissertation, Universidade de Brasília.
- Martins, Marci Fileti 2004. *Descrição e Análise de Aspectos da Gramática do Guarani Mbyá*. PhD dissertation, Universidade Estadual de Campinas.
- Mauri, Caterina 2017. Building and interpreting ad hoc categories: a linguistic analysis. In Blochowiak, Joanna; Grisot, Cristina; Durrleman, Stéphanie & Laenzlinger, Christopher (eds.), *Formal models in the study of language*. Berlin: Springer. 297-326.
- Mitãmi 2005. Translation of *Le Petit Prince* by Antoine de Saint-Exupéry. Translated by De La Nieves Domínguez & Enrique Chamorro. Asunción: Espectograf.
- Mithun, Marianne 1984. The evolution of noun incorporation. *Language* 60. 847-894.
- Mosegaard-Hansen, Maj-Britt 2013. Negation in the history of French. In Willis, David; Lucas, Christopher & Breitbarth, Anne (eds.), *The history of negation in the languages of Europe and the Mediterranean*. Oxford: Oxford University Press. 51-76.
- Mushin, Ilana 1995. Epistemes in Australian languages. *Australian Journal of Linguistics* 15. 1-31.
- O'Hagan, Zachary 2011. *Proto-Omagua-Kokama: Grammatical sketch and prehistory*. MA thesis, University of California at Berkeley.
- Praça, Wilkíria Neiva 2007. *Morfossintaxe da Língua Tapirapé (Família Tupí-Guaraní)*. PhD dissertation, Universidade de Brasília.
- Priest, Anne & Priest, Perry N. 1985. *Diccionario sirionó y castellano*. Cochabamba: Instituto Lingüístico de Verano.
- Ramirez, Henri; Vegini, Valdir & de França, Maria Christina Victorino 2017. O warázu do Guaporé (tupi-guarani): primeira descrição linguística. *Liames* 17. 411-506.
- Rodrigues, Aryon Dall'Igna 1953. Morfologia do verbo tupi. *Letras* 1. 121-152.
- Rodrigues, Aryon Dall'Igna & Cabral, Ana Suelly Arruda Câmara 2012. Tupían. In Campbell, Lyle & Grondona, Verónica (eds.), *The indigenous languages of*

- South America. *A comprehensive guide*. Berlin: Mouton de Gruyter. 495-574.
- Romano, Santiago; Cattunar, Hermán & Giannecchini, Doroteo 1916. *Diccionario Chiriguano-Español y Español-Chiriguano*. Tarija: Apostólico Colegio de María de los Angeles.
- Rose, Françoise 2011. *Grammaire de l'Émérillon Teko: Langue Tupi-Guarani de Guyane Française*. Leuven: Peeters.
- Rose, Françoise 2013. Finitization. A shift of dependency-coding strategy from Proto-Tupi-Guarani to Emérillon. *Diachronica* 30. 27-60.
- Rose, Françoise 2018. L'incorporation nominale en émérillon: une approche lexicale et discursive. *Amerindia* 31. 87-112.
- Ruíz de Montoya, Antonio 1724 [1639]. *Arte de la lengua guaraní (Notas con los escolios, anotaciones y apendices del P. Paulo Restivo, sacados de los papeles del P. Simon Bandini)*. Santa María la Mayor, Paraguay.
- Schachter, Paul & Shopen, Timothy 2007. Part-of-speech systems. In Shopen, Timothy (ed.), *Language typology and syntactic description. 2nd edition. Vol. 1: Clause structure*. Cambridge: Cambridge University Press. 1-60.
- Schleicher, Charles O. 1998. *Comparative and Internal Reconstruction of the Tupi-Guarani Languages Family*. PhD dissertation, University of Wisconsin.
- Seki, Lucy 2000. *Gramática Kamaiurá: Língua Tupi-Guarani do Alto Xingu*. Campinas: Editora Unicamp.
- Seki, Lucy & Brandon, Frank R. 2007. Uma reconstrução parcial do sistema interrogativo Tupí. In Cabral, Ana Suelly Arruda Câmara & Rodrigues, Aryon Dall'Igna (eds.), *Línguas e culturas tupí*. Campinas: Curt Nimuendaju. 259-275.
- Solano, Eliete de Jesus Bararua 2009. *Descrição gramatical da língua Araweté*. PhD dissertation, Universidade de Brasília.
- Teleman, Ulf; Hellberg, Staffan & Andersson, Erik 1999. *Svenska Akademiens grammatik 4. Satser och meningar*. Stockholm: Norstedts.
- Traugott, Elizabeth & Trousdale, Graeme 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Vallejos, Rosa 2016. *A grammar of Kukama-Kukamiria*. (Brill's Studies in the Indigenous Languages of South America, 13). Leiden: Brill.
- Van Alsenoy, Lauren 2014. *A new typology of indefinite pronouns, with a focus on negative indefinites*. PhD dissertation, University of Antwerp.
- van der Auwera, Johan & de Lissers, Tamirand Nnena 2019. Negative concord in Jamaican. *Ampersand* 6. <doi.org/10.1016/j.amper.2019.100051 >
- van der Auwera, Johan & Krasnoukhova, Olga 2020. Standard negation in Chibchan. *LIAMES: Línguas Indígenas Americanas* 20. 1-27.
- van der Auwera, Johan; Krasnoukhova, Olga & Vossen, Frens 2021. Intertwining the negative cycles. In Hamari, Arja & Veselinova, Ljuba (eds.), *The Negative Existential Cycle from a historical-comparative perspective*. Berlin: Language Science Press. 1-41.
- van der Auwera, Johan & Sahoo, Kalyanamalini 2020. *Such* similatives: a cross-linguistic reconnaissance. *Language Sciences* 81. <doi.org/10.1016/j.langsci.2018.12.002 >.
- van der Auwera, Johan & Van Alsenoy, Lauren 2016. On the typology of negative concord. *Studies in Language* 40. 473-512.
- van der Auwera, Johan & Van Alsenoy, Lauren 2018. More ado about *nothing*:

- on the typology of negative indefinites. In Turner, Ken & Horn, Laurence (eds.), *Pragmatics, truth and underspecification*. Leiden: Brill. 107-146.
- Van Gijn, Rik; Galucio, Ana Vilacy & Nogueira, Antonia Fernanda 2015. Subordination strategies in Tupian Languages. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10. 297-324.
- Veselinova, Ljuba 2013. Negative existentials: a cross-linguistic study. *Italian Journal of Linguistics* 25. 107-145.
- Veselinova, Ljuba 2014. The negative existential cycle revisited. *Linguistics* 52. 1327-1369.
- Villafañe, Lucrecia 2004. *Gramática Yuki: Lengua Tupí-Guaraní de Bolivia*. PhD dissertation, Katholieke Universiteit Nijmegen.