Chapter 1

One more thing 'thing' can do in Tupí-Guaraní languages: The Teko filler

Françoise Rose
CNRS & Université Lumière Lyon 2 & Laboratoire Dynamique du Langage

This paper proposes a first step in the study of *maʔe* 'thing' as a filler in the Tupí-Guaraní languages, in the hopes that others will follow. It offers a first detailed analysis of the discourse uses of the reflex of *maʔe* in a Tupí-Guaraní language, namely the *baʔe* noun in Teko (emer 1243). This study is based on a corpus of spontaneous texts (>5000 words) collected in French Guiana. After a description of *baʔe* as a general noun referring to non-humans and as a verb meaning 'do', the paper further describes an additional function of *baʔe*, namely as a hesitator (without morphology) for nouns and less often for other parts of speech. Prosodically, its realization as a hesitator differs from that as a noun in the average duration of its final vowel, the preceding pause, and the next word. The paper argues that *baʔe* as a hesitator is an element distinct from a noun. In a further step, three discourse uses of the noun *baʔe* are investigated for the first time in the literature on Tupí-Guaraní languages: i) as a general extender, ii) as a nominal placeholder, iii) in a rhetorical construction before a reference to a non-human being. A theoretical outcome of this paper is the extention of the map of grammaticalization paths of the Tupí-Guaraní general noun *maʔe* sketched by **AuweraAuwera2021** with paths of pragmaticalization towards disfluency markers and other types of discourse markers.

Keywords: filler, hesitator, placeholder, general extender, general noun, disfluency

1 Introduction

1.1 *maʔe* and fillers in Tupí-Guaraní languages

The noun for 'thing' is often commented on in the description of individual Tupí-Guaraní languages. Importantly, the noun *maʔe* 'thing' has been reconstructed

for Proto-Tupí-Guaraní (**Rodrigues1984**; **Jensen1998**; **Mello2000**). Two papers focus on the uses of the noun 'thing' and its grammaticalized versions in the Tupí-Guaraní language group: **Dietrich1994** and **AuweraAuwera2021**. A list of these uses follows, and most of them will later be discussed and illustrated with Teko data (Section 2).

- General term for any non-human entity

- Incorporated in a verb with an intransitivizing effect

- Used in *baʔe*-N compounds for general, nondetermined, and unpossessed entities

- Interrogative ('what', etc.)

- Negative (pro)noun ('nothing, anything'); standard negation, privative

- Nominalizer (either clause-final, or in the first position of a compound)

- Discourse markers (protest, surprise, 'yes', 'how about', 'maybe/for example', general extender 'and stuff', 'whatchamacallit')

This last use of the reflex of *maʔe* as a filler ('whatchamacallit') and fillers in general have not been a topic in the Tupí-Guaraní literature. There are only very brief and superficial mentions of fillers in the literature, and they have not yet been discussed at the group level (**Jensen1999**; **Rose2023**). Nevertheless, Dmitry Gerasimov has pointed out that the noun for 'thing' with some additional morphology *maʔe-rã* has a "whatchamacallit use" in Paraguayan Guaraní, i.e. a use as a filler. This information is only briefly mentioned in a paper dedicated to the multiple grammaticalization paths of the Tupí-Guaraní word *maʔe* 'thing' (**AuweraAuwera2021**).

The filler use of the reflex of *maʔe* has also already been reported in the first reference grammar of Teko (**Rose2011**), and is the topic of the present study, which is inspired by the literature on the richness of reflexes of *maʔe* within the Tupí-Guaraní group, as specifically noted:

> The discourse uses [of Tupí-Guaraní 'thing'] also need more work – they are mentioned for a few languages, but chances are that they have stayed under the radar for others. (**AuweraAuwera2021**)

1.2 Fillers in the Teko language

The Teko community consists of around 400 people living in two areas in French Guiana: next to the Maroni river (at the border with Suriname) and at the Oyapock-Camopi confluence (at the border with Brazil). The community (formerly known as Emerillon) is composed of the aggregation of surviving members of different ethnic groups, mainly of Tupí-Guaraní origin (**Navet1994**). The Teko language is still actively spoken and passed on to children as a native language, but it must nevertheless be considered endangered given the small number of speakers and the increasing intensity of contacts with French and Guianese Creole speakers.

Teko belongs to the Mawetí-Guaraní group (more precisely, to its Tupí-Guaraní sub-group) of the Tupí stock (**Rodrigues1984**; **roseNodate**). Tupí languages are spoken throughout Brazil, in northern Argentina, Paraguay, Bolivia, and French Guiana.

The Teko language was first described at the turn of the millennium (**Maurel1998**), and a first reference grammar was published a decade ago (**Rose2011**). In this grammar, a paragraph on fillers occurs in a section focusing on the fact that the objects of transitive verbs must always be retrievable (**Rose2011**). When an object is not specific (and is inanimate), it is expressed overtly with the general noun *baʔe* 'thing'. This is exemplified in the grammar as (1).

(1) *o-iɾuɾ o-mumuɲ=o o-wote-te **baʔe***
     3-bring 3-cook=CONT 3-alone-RED **thing**
     'She brings back and cooks **something** by herself.'   [07.035]

It is further argued in the grammar that the general meaning of this noun enables it to mark hesitation, or to facilitate a more specific rewording, as illustrated in (2) and (3), respectively.

(2) *kob (0.3) pitaŋ-am (2.3) **baʔe** (1.2) kɨto-ɾ-ehe e-iba*
     EXIST   child-TRANSF   **HES**   frog-RELN-with 3-pet
     'There is a child with his **um**... pet frog.'   [13.001]

(3) *koɾ kɨto-ɾ-aʔɨɾ (0.8) o-iɲuŋ (0.7) **baʔe-pope-dʒi** (2.4)*
     then frog-RELN-son   3-put **thing-in-LOC**
     bokaɾ-a-pe-dʒi o-iɲuŋ
     jar(Fr.)-REF-to-LOC 3-put
     'Then he puts the little frog **in something**, in a jar.'   [13.003]

In the terminology of the current volume, this means that Teko *baʔe* is used as a hesitator and a placeholder. Among "fillers", i.e. non-silence devices typically used in situations of word formulation trouble (among other functions), "placeholders" are referential terms occupying a syntactic slot in the utterance, while "hesitators" (or hesitatives) are non-referential terms that do not occupy a syntactic slot (**Podlesskaya2010**; **HayashiYoon2006**). *Thingy* is a prototypical example of a placeholder in English (**Pertejo2015**), and *uh* of a hesitator (**Tree2002**).

This preliminary description of these particular uses of *baʔe* relied on my intuition on discourse usage, translations and some morphosyntactic cues. The present study attempts to improve this description, through a corpus-based investigation of all relevant occurrences and a systematic prosodic analysis of those occurrences. For that purpose, a finer transcription of the disfluencies in the texts has been realized, thus raising the number of fillers in the corpus. The current study also benefited from the previous research on fillers, grammaticalization/pragmaticalization theory, and a review of the literature on Tupí-Guaraní languages.

1.3 Goals of the paper

The immediate goal of the present paper is to offer a first description of the discourse uses of *⃰maʔe* in a Tupí-Guaraní language, namely Teko. The main challenge is to clearly distinguish the hesitator *baʔe* from the noun *baʔe*. In order to achieve this, the morphosyntactic, semantic, discourse, and prosodic environment of *baʔe* is systematically investigated in a corpus of transcribed audio recordings.

One of the long-term goals of this study is to encourage further studies on fillers and discourse uses of 'thing' in other Tupí-Guaraní languages. Importantly, a comparative lexicon of the language group shows that reflexes of *⃰maʔe* are found in almost all languages within the branch (**chousouNodate**). There are, therefore, potentially dozens of languages for which similar studies could be launched.

A first review of the literature has revealed that information on the filler *⃰maʔe* can be found in dictionaries (e.g. **Grenand1989**: 267, on Wayampi), grammatical descriptions (e.g. **Godoy2020**, on Ka'apor) and corpora (e.g. **Gasparini2015**, on Siriono). Although the information is sparse and shallow, that makes the topic of fillers worth to be investigated within the language group.

The second long-term goal is to contribute to the expansion of the map of the historical evolution of *⃰maʔe* sketched by **AuweraAuwera2021** at the group-level, by including the results of potential pragmaticalization paths.

1.4 Methodology

The 103 audio files together with their Praat text grids, the coding sheet named *baʔe_database* and the Praat script used for the present study and presented in this section are accessible at https://hdl.handle.net/11403/teko-bae/v1.

This work is based on a rather small corpus of audio-recorded texts. Nineteen texts were collected by myself between 1999 and 2004 in French Guiana, mainly in the village of Camopi, but also in Cayenne and its surroundings.[1] Two additional texts were collected in Cayenne by Alexis Michaud in 1998. These twenty-one recordings amount to about 100 minutes and 5,352 words and make up the "oral corpus". The Teko toolbox project (**Rose2018**) as a whole also comprises written texts and a lexicon. The "written corpus" is made up of 12 published written versions of texts of oral tradition (**Renault-LescureRenault-Lescure1987**; **Ti'iwanTi'iwan1993**; **Maurel1991**; **Maurel2000**) and one leaflet introducing the euro currency (**AssociationSolidaritéGuyane2000**). This text corpus amounts to 3,793 words.[2] The lexicon comprises 1465 entries, each with parts-of-speech information, and translation in French and English. All the texts are transcribed in **ELAN2022**, translated into French and English, and annotated (with parts-of-speech and translation) at the morpheme level with Toolbox (https://software.sil.org/toolbox/). Two texts have been made public in the AILLA (http://www.ailla.utexas.org/) and Ortolang (https://hdl.handle.net/11403/sldr000870/v1) repositories.

The first step in this corpus-based study was to refine the transcription of the texts. The first round of transcription was made by myself (when still a language learner) in the field with the help of consultants who were repeating the recorded utterances. Remarkably, most occurrences of *baʔe* were transcribed, even in their use in disfluencies. So many years after conducting the transcription, it is difficult to firmly state whether the non-omission of hesitators in the transcription is due to the consultants repeating them (considering them worthy of being transcribed, contrarily to false starts or interjections) or me identifying them clearly due to their phonemic salience. Nevertheless, when some student interns and myself carefully listened to the recordings again we were able to identify a few more occurrences of *baʔe*, essentially between utterances.

---

[1]Four more text were recorded, but the recordings have been damaged after the transcriptions were realized.

[2]When cited as examples in this paper, sentences from the corpus are followed by the text number and the sentence number, separated by a dot, as in 04.035 for sentence #35 in text #4. When they are extracted from the written corpus, "w" follows the code. When they are from the audio corpus, but the audio is not available anymore, "na" follows the code.

The second step of the study was to survey the occurrences of *baʔe* in the audio-recorded texts. Then the audio for each of the 103 tokens was extracted (excluding the 95 occurrences as a verb), and the tokens were coded. The result was 27 occurrences of *baʔe* as a regular noun, 57 as a hesitator, and 19 with other uses (placeholder, general extender, rhetorical use, to be discussed in Section 4). A dataset was created (https://hdl.handle.net/11403/teko-bae/v1), with 10 pieces of information for each of these 103 occurrences of *baʔe*. First, each occurrence was attributed a unique identifier. A secondary information is the name for the accompanying audio file. The name is made up of the "type' (see below) and a code for their text and sentence number (see footnote 2), as for example "hesitator13001". Then, each occurrence was manually coded for "type", i.e. "thing", "hesitator", "placeholder", "extender", "rhetoric", or "unclear". The codes in the next three columns signal whether the item carries any morphology, whether this is preposed, or whether this is postposed. The two following columns give the form of the word following each occurrence and the number of phonemes in that word. The next three columns were filled when *baʔe* is labelled as "hesitator", "placeholder" or "rhetoric": they specify information regarding the delayed constituent, namely its part of speech, whether it is a loanword, and whether its referent is human or non-human.

The third step of the study was the analysis of some prosodic features in Praat (**BoersmaBoersma2023**) and is illustrated in Figure 1. As mentioned above, each occurrence of *baʔe* was extracted as a short audio file from the text recording, along with the preceding and the following words (and pauses if present). In each file, each word and pause was segmented, as well as the vowels of *baʔe*.[3] In the first tier, the vowels /a/ and /e/ were annotated. In the second tier, the interval corresponding to the form *baʔe* (without possible morphology) is coded with "thing" when *baʔe* is used as a plain noun, with "extender", "placeholder" or "rhetoric" when this noun has special discourse uses (see Section 4), with "hesitator" when it is used as a hesitator, and with "unclear" for the remaining cases. On this second tier, "pw" and "fw" code the preceding and following words, and "pp" and "fp" the preceding and following pauses.[4] There may be intervals without annotation, which correspond to bound morphology on *baʔe*. The 103 audio files and their corresponding annotations are accessible at

---

[3]Note that the initial plosive may be prenasalized [mb], and that the glottal stop phoneme may be produced without a closure of the vocal folds (**Rose2021**). In this case, this phoneme is not segmented as an interval.

[4]What are considered pauses are silent pauses, but some of these silent pauses include search sounds (typically *mm*).

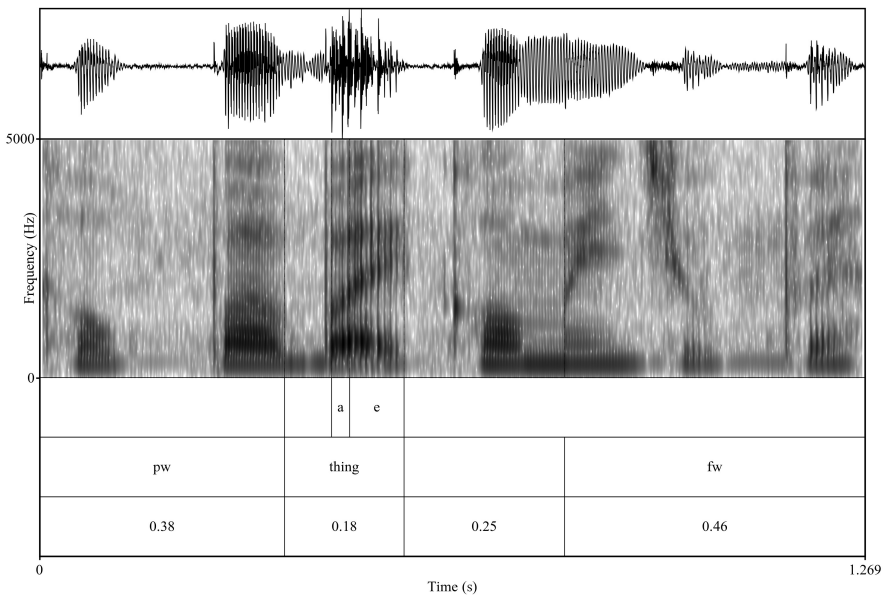https://hdl.handle.net/11403/teko-bae/v1. The Praat pictures presented in this paper contain an additional tier to make the times easily readable.

Figure 1: Praat analysis of an occurrence of *baʔe* as a noun

A script (LengthExtraction.praat, accessible at https://hdl.handle.net/11403/teko-bae/v1) extracted the duration of *baʔe*, the /a/, the /e/, the preceding and following words, and the preceding and following pauses. The results have been inserted into the *baʔe* dataset available in the supplementary materials. This enables a systematic comparison of the duration of *baʔe* and the elements that surround it depending on the analysis of its function in discourse.

In the examples, the elements under study, usually *baʔe*, are indicated in bold case, while the targets of the fillers (when their identification is clear) are displayed underlined. The duration of pauses over 0.15 seconds is given into parentheses (in seconds with deciseconds) when relevant.

1.5 Overview of the paper

Table 1 shows the number of occurrences of each use of *baʔe* in the corpus.[5] For the rare uses, a thorough search for additional examples (counted in parentheses) was conducted in the written corpus. The paper will mostly account for the frequent use of *baʔe* as a hesitator and compare it with its basic use as a noun. It will also sketch a preliminary description of rarer discourse uses of *baʔe* and provide information on its use as a verb.

Table 1: Relevant occurences of *baʔe* in the corpus.

Type	Tokens
Noun	27
Verb	95
Hesitator	57
Noun as an extender	4 (+2)
Noun as a placeholder	8
Rhetorical N+N phrase	4 (+2)

First, Section 2 will present the uses of *baʔe* in Teko. Then, Section 3 will analyze some uses of *baʔe* as a hesitator, which in synchrony cannot be analyzed as a noun anymore. Section 4 will later explore the discourse uses of the *baʔe* noun 'thing'. Finally, Section 5 will address diachronic hypotheses on the development of the different *baʔe* items and their use in Teko.

2 The uses of *baʔe* in Teko

The noun for 'thing', reconstructed as **maʔe* for Proto-Tupí-Guaraní, gave rise to a reflex in Teko, the noun *baʔe*. In line with some known paths of evolution within the family (**Dietrich1994**; **Jensen1998**; **AuweraAuwera2021**), the same form is employed in the language for nominalization (Section 2.2) and interrogation (Section 2.3), but not for negation. Beside these well-described uses of the reflexes of **maʔe*, Teko *baʔe* is also used as a verb (Section 2.4), as well as with some discourse functions, which are the topic of this paper.

Before discussing the discourse uses of *baʔe*, this section summarizes what is known about its other uses in Teko: as a noun in Section 2.1, in nominalization

---

[5]Three other occurrences are not included in Table 1, as their analysis remains unclear.

in Section 2.2, in interrogation in Section 2.3, and as a verb in Section 2.4. This background information is necessary to later determine the part of speech of the *baʔe* elements under study in Section 3 and Section 4, and to discuss diachronic paths of evolution in Section 5.

2.1 Noun

Teko *baʔe* is a noun with a general non-human meaning. This section will describe this term first on the morphosyntactic level, and second, on the semantic level.

The nominal morphosyntactic features that undoubtedly make *baʔe* a noun are the following:

- it can be the head of a noun phrase, in the position of S, as in (4), O, as in (5), or object of postposition, as in (6);

(4) *o-ho-pa*   ***baʔe-kom-a=nam***   *o-apɨg=o*   *kupa=o*
    3-go-COMPL **thing-PL-REF=when** 3-sit=CONT PL.S=CONT
    'When all the **animals** had left, they (the men) sat down.'   [23.084w]

(5) ***baʔe***   *t-a-pɨhɨg*
    thing EXH-1SG-catch
    'I want to catch **them** (girls who later happen to be mermaids).'   [32.043]

(6) *apam-a-**baʔe**-pe=ãhã*   *node-ɾ-apiʤ-a-pope*
    foreigner-REF-**thing**-about=only 1INCL-RELN-house-REF-in
    '(It is) just about foreigners' **artefacts** in our house [freezer, stove, TV and so on].'   [08.014]

- it can be determined by a demonstrative, as in (10);

- it can be modified by a quantifier, an adjective or a relative;

- it is countable, and as such can take a plural suffix, as in (4) and (7);

- it can take a possessive person prefix, as in (7), or be preceded by a noun for its possessor, as in (6), which makes it an optionally possessed noun.

(7) *kob-pa-katu*   *i-**baʔe**-kom i-wɨɾakotɨ*
    EXIST-COMPL-good 3-**thing**-PL 3-under
    'There are all her **belongings** below it.'   [20.010]

On the semantic level, *baʔe* is a hypernym for both inanimate objects, as in (7), and non-human animates, covering all animals and other non-human beings, as illustrated in (4) and (5), respectively. Consequently, the common gloss 'thing' for *baʔe* and its cognates in other Tupí-Guaraní languages is misleading. A better translation is 'non-human'. In this paper, 'thing' is taken as a proxy for non-human in order to follow the literature on Tupí-Guaraní languages.

(8) o-ho-pa   **baʔe**-*kom-a=nam*   *o-apɨg=o*   *kupa=o*
     3-go-COMPL **thing**-PL-REF=when 3-sit=CONT PL.S=CONT
     'When all the **animals** had left, they (the men) sat down.'   [23.084w]

(9) **baʔe** *t-a-pɨhɨg*
     thing EXH-1SG-catch
     'I want to catch **them** (girls who later happen to be mermaids).'   [32.043]

There is only one example in the corpus that refers to a human referent, given in (10). Maybe the referent, a white girl, is culturally categorized as *baʔe*. In fact, the most basic noun referring to human beings is *teko,* and in a restricted sense, it can refer to members of the Teko-speaking community only (**CachineCachine2020**), of which the 'white girl' is obviously not a part.[6] Alternatively, the use of *baʔe* with a human referent in this sentence could be explained through discourse uses of *baʔe* (see Section 4.3).

(10) si-ɾo-nan   *aŋ* **baʔe** *panaɪ̃sɪ̃*   *wãɪ̃wɨ-am*
     1INCL-SOC.CAUS-run DEM **thing** White_person woman-TRANSF
     'We have kidnapped this (thing) white girl.'   [10.030na]

The noun *baʔe* must be considered a general noun, i.e. a superordinate term that refers to a broad category of entities (**Cutting2019**). General nouns show a high frequency of use, a weak semantic load, and a wide array of potential referents, and one must consequently rely on the previous or subsequent co-text, or on the exophoric context to identify their referent (**HallidayHalliday2013**; **AdlerAdler2018**). General nouns participate in vague expression, i.e. language that is "inherently and intentionally imprecise" and has "lexical and grammatical surface features [...] that may refer either to specific entities or to nothing in particular" (**Cutting2007**). And indeed, tokens of the noun *baʔe* in spontaneous

---

[6]In contrast, there are the nouns *apam* for 'foreigner, non-Teko' and *karai* for 'Brazilian', as well as loanwords like *panaɪ̃sɪ̃* (from French) for 'French' or 'white person other than Brazilian', and *mun* (from Guyanese Creole) as a general term for 'person'.

speech vary in having specific referents (identifiable or not) or non-specific referents.

In (7) to (9), which are devoid of any disfluency, the referents of the noun *baʔe* are specific and identifiable, either anaphorically, cataphorically or exophorically. In (7), from a picture description task, the speaker makes reference to a boucan[7] and pets in the previous sentence, and to firewood in the following one: all these elements are considered to be individually related to the housewife seen in the picture. Example (8) follows a dozen of sentences in which a number of animals (snake, mosquito, jaguar, agami, spider monkey and howler monkey) are introduced one after the other. *Baʔekom* refers back to the globality of these individual referents. Example (9) is from a text with two female heroes to which *baʔe* refers.

In other cases, the noun *baʔe* encodes a referent which is specific but either not clearly identifiable or not identifiable at all. In (11), the referent of *baʔe* is vague for the hearer: it could refer to elements of the landscape, artefacts, etc. And when *baʔe* is the questioned constituent in open questions, it refers to specific but indeterminate non-human(s) (see Section 2.3 about the use of *baʔe* in questions).

(11) *"waɾeʔete-aʔu   a-maʔẽ **baʔe**-kom-a-ɾ-ehe   pe-ɾ-upi*
      be_beautiful-TOP_SW 1SG-see **thing**-PL-REF-RELN-to path-RELN-along
      a-ho-nam",   eʔi-ra
      1SG-go-when 3.say-PFV
      ' "I have seen nice things on the way", she said.' [speaker is relating
      linguist's account of a trip to Surinam]   [03.002]

Still in other cases, the noun *baʔe* encodes a non-specific referent. This occurs in particular in two constructions where a nominal term is required by the morphosyntax and can be filled with *baʔe*, the referent of which is then non-specific. The first construction involves a transitive verb with a generic object, as illustrated in (12) and discussed in **Rose2011**. Even though the speaker does not feel the need to specify the patient, the syntax of Teko requires that an object be retrievable.[8] The object position is then filled in by general terms like *baʔe*, or also *baʔezaʔu* 'food'[9] after verbs of cooking and ingestion.

(12) *zawapinim=ãhã **baʔe** o-ʔu=o*
      jaguar=only   **thing** 3-eat=CONT
      'Only the jaguar is eating (something).'   [29.031w]

---

[7]A 'boucan' is a wooden frame on which meat or fish is smoked.

[8]See also discussion and (1) in Section 1.2: it is precisely in the section of the grammar (**Rose2011**) which explains this syntactic requirement that the issue of fillers was first mentioned.

[9]The nominal stem *baʔe-za-ʔu* 'food' is made of *baʔe* 'thing', *za*, which is formally similar to the indeterminate subject prefix, and the verb *ʔu* 'ingest'.

The second construction is a Noun-Noun construction involving an obligatorily possessed noun in the head (N2) position, with *baʔe* in N1 position expressing a non-specific non-human possessor, as exemplified in (13) and (14).[10]

(13)   a.   *kunami-potɨɾ*
           kunami-flower
           'kunami flower (type of creeper)'

        b.   *baʔe-potɨɾ*
           thing-flower
           'flower (general term)'

(14)   a.   *basakaɾa-ɾ-upiʔa*
           hen-RELN-egg
           'chicken egg'

        b.   *baʔe-ɾ-upiʔa*
           thing-RELN-egg
           'egg (general term)'

In these two constructions, *baʔe* appears bare in the syntactically required position, and is either translated as 'something', or not translated at all in French or English. These are languages in which objects of transitive verbs are not obligatorily overt and where nouns are never obligatorily possessed, so that these positions can be left empty.[11] Even though the general noun here substitutes for a more specific noun like placeholders do, it is vague but does not additionally invite the hearer to replace it with a specific noun based on contextual knowledge. It must be noted that, "in addition to their highly general semantic content, placeholders signal that the speaker is not able or willing to provide a more specific target expression which would be necessary to convey the message. Instead, the placeholders communicate that their semantic content needs to be enriched by context knowledge" (**HenneckeHennecke2022**). I therefore do not analyse *baʔe* in (12) to (14) as a placeholder. Its actual placeholder uses will be discussed in Section 4.2.

---

[10]See **Rose2011** on Teko obligatorily possessed nouns and their person prefix *zo-* for non-specific human possessors.

[11]Nevertheless, "in English there is a tendency to fill up all the relevant valency positions even if there is no definite entity participating in the action/state expressed by the verb. *Things* is one of the dummy objects which help to fulfil this function." (**Fronek1982**)

2.2 Nominalization

Nominalization can only be considered an unproductive function of *baʔe.*, which is found in some very rare nominalizations without other nominalizing devices, such as the compound in (15).[12]

(15) *baʔe-kʷa-wəɾ*
       thing-pass-FREQ
       'beads string reminding the order of traditional chants'

[CachineCachine2020:15]

Incidentally, the productive Teko relativizer/complementizer =*mãʔẽ* (**Rose2011**) has been reconstructed as a clause nominalizer **βaʔe* in Proto-Tupí-Guaraní (**Jensen1998**), and this reconstruction is hypothesized to be linked to **maʔe* 'thing' by **AuweraAuwera2021**. This morpheme is added to the right of finite clauses, and the resulting phrase fills the nominal position of an object of transitive verbs or a modifier of a head noun to its left, such as *baʔekʷəɾ* in (16).

(16) *aŋ baʔekʷəɾ[13] [a-mebeʔu-taɾ-a=**mãʔẽ**].*
       DEM story   1SG-tell-FUT-REF=REL
       'This is the story that I am going to tell.'   [01.001]

2.3 Interrogation

Questions in Teko are marked with an interrogative clitic that is either added as an enclitic to the initial constituent, as in (17), or as a proclitic to the second constituent, as in (18). The clitic for wh-questions is *to/ta*, the one for polar questions is *so/sa*, and the one for exclamative interrogation is *sipo*. The initial constituent is under the scope of the interrogation, and there is no additional morphosyntactic modification to the clause.

(17) ***baʔe=to** pe-(e)kaɾ=iɲ, e-paɾɨ-kom?*
       thing=INTER 2PL-search=CONT 1SG-grandchild-PL
       'What are you looking for, my grandchildren?'   [12.114na]

---

[12]The other cases are *baʔe-wa* 'restaurant' where *wa* is a non-productive verbal root for 'eat', *baʔe-ɾ-aɨ* 'illness' where *aɨ* 'be in pain' is not a verb but a type of adjectival predicate (called "attributive" in **Rose2011**), and *baʔe-za-ʔu* 'food' (see footnote 9).

[13]This term results from the lexicalization of *baʔe* 'non-human' and the suffix *-kʷəɾ*, which is often translated in English with the prefix 'ex-'.

(18) **baʔe=ne=eʔe**   **ta=aŋ**   o-baʔe? awa ta=aŋ   o-ɨg?
     thing=CONTR=INTENS **INTER**=DEM 3-do   who INTER=DEM 3-arrive
     'Who (lit. what) did this [about cooking]? Who arrived?'   [22.186]

Table 2 lists Teko interrogative forms. Only the first one is a single morpheme dedicated to interrogation. All others are either based on *baʔe* or on *ma*, which could be analyzed as an indefinite root. **AuweraAuwera2021** actually relate the latter root to *maʔe.

Table 2: Interrogative forms and their likely sources

interrogative form	gloss	source	source gloss
awa	'who?'		
baʔe	'what?'		'thing'
manan ~ manani	'how?'	*ma+nan*	Q+'so'
manam	'when?'	*ma+nam*	Q+'when'
matɨ	'where to / from?'	*ma+tɨ*	Q+towards
mananãhã	'how many?'	*ma+nan+ãhã*	Q+'so'+'only'
maŋ	'which?'	*ma+aŋ*	Q+DEM
baʔamõ	'what for?'	*baʔe+am(õ)*	'thing'+TRANSF
baʔe-ɾ=ehe	'why?'	*baʔe-ɾ=ehe*	'thing'+RELN+POSTP
baʔe=wi	'from what?'	*baʔe=wi*	'thing'+ABL

The relevant research question here is whether the *baʔe* form used in questions is an interrogative pronoun or whether it can still be considered as the noun *baʔe* 'thing'. In the latter case, interrogation would still be overtly encoded by the interrogative clitic and prosody.

Morphosyntactically, *baʔe* in questions seems to behave still as a noun: it is the head of a noun phrase, as it can be the object of postpositions like *ehe* or *wi*. Nevertheless, no example case could be found in that context with a demonstrative, a plural suffix or a possessor, which are listed as nominal features of *baʔe* 'thing' in Section 2.1.

Semantically, *baʔe* in questions is used for non-human participants (translated as 'what' in the dictionary (**CachineCachine2020**)). There is nevertheless some ambiguity on whether it could have a wider semantic scope in questions. There are two occurrences in traditional narratives where the questioned referent would be expected to be human in normal situations (given the situation, i.e. someone cooking in (18) and someone having sex with a girl in (19)). They are both translated by 'who'. Moreover, in (18), *baʔe* is later reworded with *awa* 'who', a question word specific to human participants.

(19)   **ba***ʔe=sipo*   *zewe e-me-menõ?*
      thing=INTER/EXCL daily 1SG-RED-copulate_with
      'Who sleeps with me every night?"   [01.006]

However, in these two cases, the identity of the targeted character is neither straightforwardly human or non-human: it changes within the narrative (and the speakers know that in advance). In (18), the being who cooks in the house of a single man when he goes out is a woman dressed up as a macaque, and in (19), the incestuous boy who imposes sexual relationships on his sister at night ends up being chased far into the sky and turns into the moon. It is therefore difficult to decide, on the basis of the existing corpus, whether the use of interrogative *baʔe* for apparently human participants is a semantic shift, which could be evidence of grammaticalization of the noun into an interrogative pronoun, whether these two occurrences of *baʔe* reflect the non-human part of the targeted participant, or else whether *baʔe* is used to leave the questioned entity unspecified as human or non-human. With my current knowledge of the language, I would rather consider *baʔe* in wh-questions to be still a noun, with a non-human referent.

2.4 Verb

The form *baʔe* is also a verb in Teko. It should be noted that nouns and verbs are easy to distinguish in this language because of their combination with different person prefixes (**Rose2011**). The transitive verb *baʔe* means 'make, do'. Its main meaning ('make') is 'fabrication', as in (20), but it can also express creation without manipulation, as in (21).

(20)   *beɟu*   *(0.3) o-***baʔe** *(0.3) zapẽhẽ-pope*
      cassava_bread   3-**make**   griddle-on
      'She makes the cassava bread on the griddle.'   [06.007]

(21)   *lekol*   *a-***baʔe**
      school 1SG-**make**
      'I made the school [says the mayor, who took the decision to have a school in the village].'   [15.049]

The verb *baʔe* is also translated as 'do', i.e. with a meaning not encompassing the 'creation' component, as illustrated in (22). The action it refers to can be deictically inferred, either exophorically in the speech situation (23) or anaphorically in the text (24). Furthermore, *baʔe* can serve as a manner demonstrative verb (**Guérin2015**) when introducing ideophones, which depict events, i.e. actions or states, in Teko (**Rose2024**). This light verb use is illustrated in (25).

(22) *aipo=po   a-baʔe-taɾ*
      now=INTER 1SG-do-FUT
      'Will I do something now?'   [01.007]

(23) *ani (1.4), ani, mamen nan   eɾe-baʔe*
      no   no   NEG.IMP MAN.DEM 2SG-do
      'No, no, don't do that (lit. like this).'   [21.084]

(24) *a-baʔe-ta-kom   a-ʔe-ɾa*
      1SG-do-FUT-PL.O 1SG-say-PFV
      'I told you that I will do (it to) them.'   [21.205]

(25) *moŋ   o-**baʔe***
      IDEO.be_dark 3-**make**
      'S/he turns the light off.'   [elicited]

When introducing an ideophone, *baʔe* contrasts with the partially synony-
mous verb *ʔe* 'do, say', and is almost always used with an animate subject. It
is in fact used when the subject participant exerts some control over the state of
affairs expressed by the ideophone, as in (25), while *ʔe* does not imply control, as
in (26).

(26) *moŋ   eʔi*
      IDEO.be_dark 3.do
      'It is dark.'   [elicited]

With the latter interpretations, the verb *baʔe* is a general verb, i.e. it refers to
a wide category of actions.

None of the corpus examples involving *baʔe* as a verb involve disfluency cues,
and none show a subsequent verb that would elaborate on a more specific action:
baʔe as a verb cannot be considered to be a placeholder in the corpus.

Interestingly, the Proto-Tupí-Guaraní reconstruction **maʔe* is only glossed as
a noun (and as an interrogative, by Rodrigues). The reconstruction for 'do, make'
is **apo* (**Rodrigues1984**: 36; **Mello2000**: 235-236). **Mello2000** indicates that only
in Teko and Kaapor has **apo* been substituted by a reflex of **maʔe*.

3 *Baʔe* as a hesitator

In many of its occurrences (57 in the audio corpus), it is quite clear that *baʔe* is
linked to disfluency. This section will present the reasons why it should not be

considered as the noun 'thing' in these cases, but as a different word analyzable as a hesitator, as in (2). A hesitator is a marker of hesitation which differs from a placeholder in that it "is not produced as a syntactic constituent of an utterance-in-progress and therefore does not occupy any specific syntactic slot within the structure of an unfolding utterance" (**HayashiYoon2006**: 507). Examples of hesitators are *um* and *uh* in English (Clark & Fox **Tree2002**), and *este* in Amazonian Spanish (**VallejosYopán2023**). The number of 57 tokens of hesitators in the 5,352 word corpus results in an average frequency of 11 hesitators per thousand words.

The underlying reason for considering that these occurrences of *baʔe* cannot be nouns is their function: they are not referring, but rather "gaining time" in the process of searching for a word or a whole formulation. Actually, they are never translated in the free translation offered by consultants. Their analysis in terms of hesitator is based upon four formal criteria, which are discussed in this section.

The first criterion is the syntactic distribution. These occurrences of *baʔe* do not fulfill the syntactic positions typical of nouns, but instead come in a variety of syntactic contexts. In particular, in comparison to the discourse uses of *baʔe* to be presented in Section 4 (extender, placeholder, and rhetorical use), they are not restricted to positions that could be interpreted as nominal positions. They are, for example, never found as an object of a postposition. Table 3 shows the variety of elements that follow these *baʔe*, i.e. the "delayed constituents" in the terminology of disfluency analysis.

Table 3: Type of delayed constituent following *baʔe* as a hesitator

Delayed constituent	Tokens in the audio corpus	Example
noun phrase	40	(27)
verb	9	(28)
other category / clause	8	(29)
total	57	

Constituents delayed with *baʔe* are most often noun phrases (some within a postpositional phrase), but also sometimes verbs or other categories. In the latter case, *baʔe* appears in clause-initial position. Examples follow in (27) to (29).

(27) *nan-am-a=ko=ɲ*   *mamã* **baʔe** *(2)* *nana=ne*
      thus-TRANSF-REF=INJ=CONT mommy **HES**   pineapple=CONTR

o-boʔi-ʔi
3-split-DIM
'At that time, Mum was cutting into pieces the um... pineapple.' [21.134]

(28) *mɨn-a=we=zepe,*   ***baʔe*** *(0.7) oʔu-o-ʔu   wasei*
formerly-REF=also=CONCES **HES**   RED-3-eat açaí
'In the old days, um... they were often eating açaí.' [31.042]

(29) ***baʔe*** *mokoɲ wãĩwĩ-kom (0.9) o-iba=we   o-menõ*
HES two   woman-PL   3-pet=also 3-copulate_with
'Um... both girls have sex with their pet.' [32.011]

The second criterion is that when these occurrences of *baʔe* precede a nominal element, this is not restricted to nominals expressing non-human referents. The referents expressed by delayed nominals can be inanimate participants (27), animals (30), other non-human beings (31) or humans (32). There are notably four tokens of nouns referring to humans that follow this type of *baʔe*. This fact does not align with the semantics of the *baʔe* noun.

(30) *kob   (0.3) pitaŋ-am   (2.3)* ***baʔe*** *(1.2) kɨto-ɾ=ehe   e-iba*
EXIST   child-TRANSF   **HES**   frog-RELN=with 3-pet
'There is a child with his um... pet frog.' [13.001]

(31) ***baʔe*** *(0.7) tebesi-kotɨ (0.8) tebesi-kotɨ=ne=eʔe   o-ho-ŋ*
HES   Tebesig-at   Tebesig-at=CONTR=INTENS 3-go-PL.S
'They are going to the um... Tebesig's place [Tebesig are under-ground white beings].' [21.112]

(32) *mataʔeɾe[14]* ***baʔe*** *de-zaɾ-a-kom*
where   **HES** 2SG-master-REF-PL
'Where are um... your masters?' [22.143]

The third criterion for not considering these occurrences of *baʔe* as nouns is the absence of morphology. There are no examples with a plural suffix (even before a plural noun, as in (32)), a possessive prefix (even before an adnominal possessive construction, as in (30) or (32)), or a postposition (even before a noun which is the object of a postposition, as in (31)). This is particularly visible in (33), where the delayed noun takes a lot of morphology, but *baʔe* is bare.

---

[14]This word form is partly inaudible, therefore, doubts remain on the proper transcription.

(33) ***baʔe** (0.2) zawaɾ-a-kom-a-ɾ=ehe=ãhã=pɨʔa=teɾa   o-maʔẽ*
     HES   dog-REF-PL-REF-RELN=with=only=first=? 3-see
     'Um... only the dogs he first saw.'   [19.050]

The fourth criterion for not considering these occurrences of *baʔe* as nouns is their prosody. When used as a hesitator, *baʔe* is generally lengthened and preceded and followed by a pause. This is illustrated in Figure 2, representing (34).

(34) *kɨwo=ne*   *pe-itʃɨ-itʃɨg*   *(0.2)* ***baʔe** (0.2) zadupa t-o-tui*
     towards_here=CONTR 2PL-REP-drop   **HES**   genipa PURP-3-be
     'But it's here that you've dropped the um... genipa fruit (so that it stays there).'   [22.073]
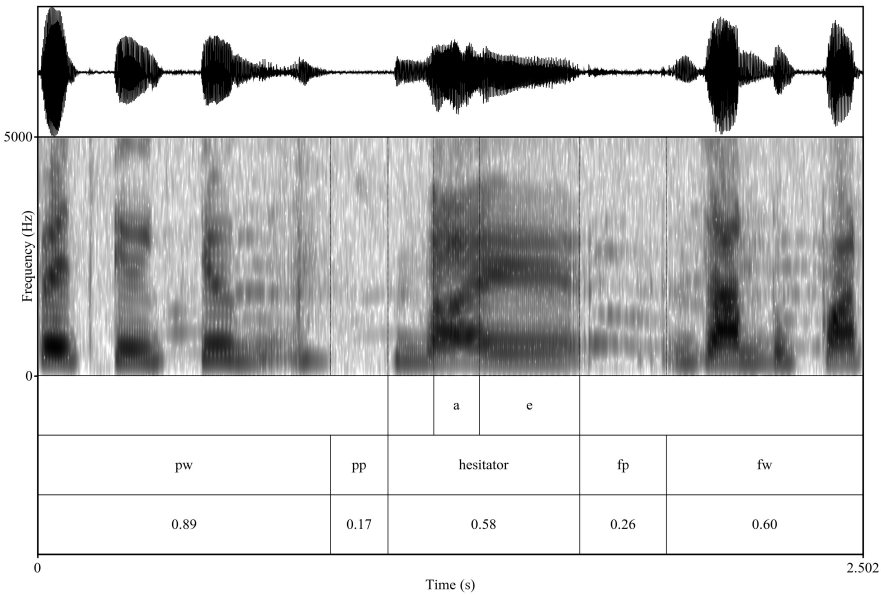
Figure 2: Praat analysis of an occurrence of *baʔe* as a hesitator

The length of *baʔe* when used as a noun (without any specific discourse use) and as a hesitator differs quite remarkably. This is mostly due to final lengthening. Table 4 compares the duration of /e/ in the basic nominal use of *baʔe* (in the 11 occurrences without any morphology) and its hesitator use.

Table 4: *baʔe* final lengthening

Functions of *baʔe* (w/out morphology)	Duration of /e/ (ms)
Noun (11)	189
Hesitator (57)	247

The hesitator is made more salient by surrounding pauses. The second and fourth columns of Table 5 show that the presence of pauses preceding and following *baʔe* is much more frequent when it is a hesitator rather than a noun. Additionally, in the cases where *baʔe* is actually followed by a pause, the pause is much longer when it follows the hesitator than when it follows the noun (last column of Table 5). In contrast, there is no substantial difference in the duration for the pause preceding *baʔe*, when there is a pause (third column of Table 5).

Table 5: Pauses before and after *baʔe*

Functions of *baʔe* (all included)	Presence of preceding pause	Duration of preceding pause (ms)	Presence of following pause	Duration of following pause (ms)
Noun (27)	48%	1205	33%	563
Hesitator (57)	77%	1230	65%	986

As a conclusion, the syntactic, morphological, semantic and prosodic characteristics of *baʔe* are sufficient to distinguish two lexical units and identify particular occurrences as either nouns or hesitators. This result is in tune with the discussion of the development of fillers out of demonstratives in several studies, where fillers are analyzed as elements distinct from the demonstratives, rather than as an additional function of them (**HayashiYoon2006**; **VallejosYopán2023**). Within the Teko linguistic system, "hesitator" could be considered a separate part of speech.

4 Discourse uses of the noun *baʔe*

This section proposes that the noun *baʔe* shows three discourse uses: as an extender (4.1), as a placeholder (4.2), and a third use, as a rhetorical device (4.3). As

shown in Table 1, these discourse uses of the noun *baʔe* are rare in the present corpus, therefore the analyses in this section are clearly preliminary, which calls for further examinations of more tokens in a larger corpus for confirmation.

In these three cases, the element under study shows the morphosyntactic properties, formal realization and the meaning of the noun *baʔe* (see Section 2.1), i.e. its part of speech has not changed. They are therefore considered discourse uses of the noun.

4.1 Extender use

In four sentences in the audio corpus (and two additional sentences in the written corpus), *baʔe* follows either one noun phrase or several noun phrases and is interpreted as extending an ad hoc category created by the preceding noun phrase(s). An ad hoc category is "created on the fly, for communicative purposes" (**MauriMauri2018**).

The use of *baʔe* as a trigger for this abstraction can be called "general extender" for five of these examples, and "specific extender" for the other one (**OverstreetOverstreet2021**).

> So-called general extenders are a group of expressions which typically exhibit a basic syntactic structure, [CONJUNCTION + NONSPECIFIC NOUN PHRASE] (e.g. *and such*, *or something*), and occur at the end of a list to indicate the existence of additional referents. (**MauriMauri2018**)

Two examples are given as illustrations:

(35)   *ikemɨn-a=nam*   *(0.5) za-siŋaɾ,*   *za-weɾa,*   *wɨwa-pe*   *piɾa*
     old_days-REF=SUB   INDET-poison_fish INDET-hunt arrow-with fish
     za-ʤika   *(0.8), di-za-moboɾ-i*   *(1.3) tramaj*   *(1.2)* **_baʔe-kom_**
     INDET-kill   NEG-INDET-throw-NEG   trammel   thing-PL
     'In the old times, one would poison fish, hunt, kill fish with arrows, one did not throw the trammel net **and this sort of things**.'   [30.020]

(36)   *dati*   *aɾakapusa (0.8), dati*   *ʃort,*   *t-iɾu,*   **_baʔe-kom_**
     EXIST.NEG gun   EXIST.NEG shorts NSP-clothes thing-PL
     'There was no gun, no shorts, no clothes, **and so on**.'   [30.012]

The ad hoc category is abstracted either from the one representative exemplar expressed by the single noun phrase, or from the list of core elements of this category expressed by the successive noun phrases. As the extender expression is

vague, the ad hoc category must be pragmatically inferred from the context. In
(35), the category "non-traditional devices used to fish" is inferred from the single
exemplar of the trammel net, but could also include metallic fish hooks. In (36),
the category "artefacts brought to us by non-indigenous people" is built on the
category members that are overtly expressed, i.e. 'gun, shorts and clothes', but im-
plicitly includes (given the context) items like 'watch, bra, TV'. In the single case
of a specific extender given in (37), the general noun of the extender construc-
tion is specified by a relative clause, a structural option known for inserting more
specific information about the ad hoc category (**OverstreetOverstreet2021**).

(37) *o-pɨhɨg tɨkadɨɾ   o-ʔoɾam idʒa   baʔe ai-mã̃ʔẽ-kom*
      3-take bullet_ant 3-with   various thing hurt-REL-PL

      'he (the shaman) takes the bullet ants (and) with it various things that
      hurt (i.e. other insects to apply on bodies)'   [28.014w]

Three of the five occurrences of the general extender were not initially trans-
lated by consultants. A fourth one is translated by the French placeholder *les
trucs* (~ 'the stuff') and the fifth by a phrase expressing a vague category *ce genre
de choses* ('this sort of things'). The absence of translation in three cases may
be explained by two different reasons which are not exclusive: first, consultants
are not trained translators and have a limited command of French, which proba-
bly does not cover expressions like *et cetera*; second, it could be that they regard
these optional linguistic elements as little informative, as in the case of hesitators
described above.

In the occurrences of *baʔe* as an extender, it still clearly shows its basic nominal
features:

- It is the head of a noun phrase, the same as the preceding noun phrases,
  namely the object of the verb in (35) and the argument of the existential
  copula in (36).

- It can take the plural marker *-kom*. It does so in four of the six examples,
  as in (35) and (36).

- In two cases, *baʔekom* is preceded by the modifier *idʒa* 'various', as in (37).

- In all six cases, *baʔe* can be said to maintain its non-human meaning, be
  it for inanimate items in four cases, including (35) and (36), or animals in
  two cases, including (37). There is no sign of semantic bleaching.

Since there are no phonetic, semantic or morphosyntactic signs of decategorialization, *baʔe* is seen as a noun used in a construction that serves pragmatic purposes. It should further be noted that there is no sign of disfluency in the six occurrences: no salient prosody on *baʔe* like vowel lengthening, no remarkable pause or false start, and no delayed constituent. Figure 3 illustrates the end of (36) and the beginning of the following clause, where *baʔekom* immediately follows *tɨɾu*, and is not lengthened. The long pause is an interclausal pause.
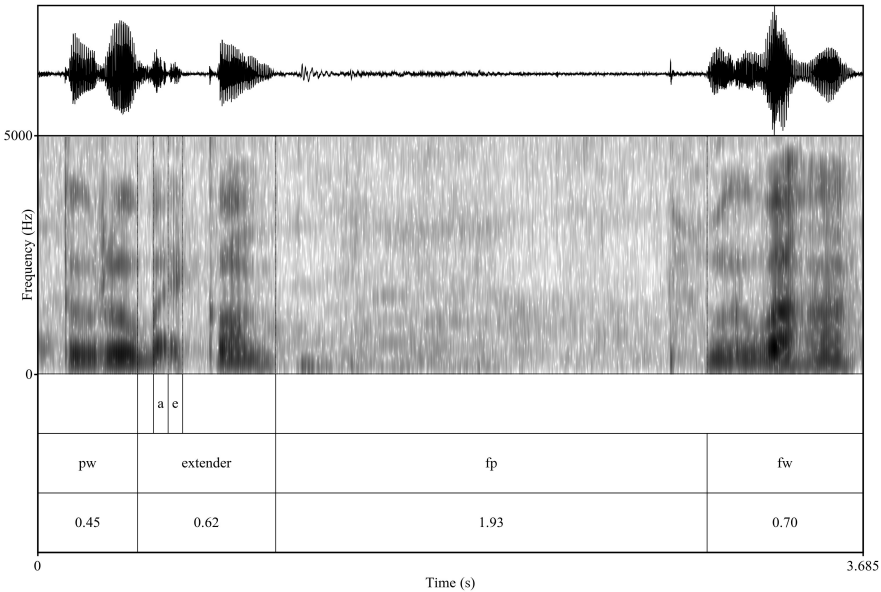
Figure 3: Praat analysis of an occurrence of *baʔe* as an extender

Even though **MauriMauri2017** identify other formal types of general extender, the construction is typically made of a connective, an indefinite or general element, and an element encoding similarity. Not all elements are necessarily found in a general extender construction. In the case of Teko, there is indeed an element with a general meaning, namely the noun *baʔe*. It has been explained that this noun might have either a specific or a non-specific referent in its basic use (see Section 2.1), but it always has a non-specific meaning in the general extender construction: vagueness systematically associates with this noun when used as an extender.

However, the Teko extender construction does depart from the prototypical

construction: it does not involve an independent connector or an independent expression of similarity. This is not surprising, as the language does not have an independent connector to link noun phrases (**Rose2011**),[15] nor nouns for 'type, gender, sort, class, category' (**CachineCachine2020**). However, in five of the six examples, the *-kom* plural marker is used. This marker, the rich and complex distribution of which is detailed in **Rose2012**, can actually be used with a connective function, as in (38), and is also used as an associative plural, as in (39).[16] Its two functions of encoding connection and plurality are possibly at play in adding possible category members to build the ad hoc category in the general extender construction, without its presence being necessary for the construction. In all cases, in the absence of a connective, the general extender construction is interpreted as adjunctive.[17]

(38) *Papa mamã-**kom**-a-pu ɾi a-ʤu=ɲ*
      Dad Mum-**PL**-**REF**-at   1SG-be=CONT

      'I am at Mum and Dad's.'   [22.027]

(39) *i-jɨ-**kom**   (0.2) "m" eʔi*
      3-mother-**PL**   yes 3.say

      'The parents (lit. her mothers / their mother) say "yes"'   [32.007]

4.2 Placeholder use

There are eight cases in the audio corpus where the noun *baʔe* takes some morphology, and is further specified by a specific noun. This pattern is illustrated in (40).

(40) *koɾ kɨto-ɾ-aʔiɾ   (0.8) o-iɲuŋ (0.7) **baʔe-pope-ʤi** (2.4)*
      then frog-**RELN**-son   3-put   **thing-in-LOC**

      bokaɾ-a-pe-ʤi   o-iɲuŋ
      jar(Fr.)-**REF**-in-**LOC** 3-put

      'Then he puts the baby frog **in a thingy**, in a jar.'   [13.003]

---

[15]The conjunction *ʔoɾam* 'with' preceded by a coreferential third person prefix *o-* might serve a function similar to that of a conjunction. It is found in the single example of a specific extender (37).

[16]The two clear examples of the associative plural use are on kin terms, which is cross-linguistically expected (**DanielMoravcsik2013**). The associates to the focal member are other family members.

[17]"There are two distinct types [of general extenders in English]: those beginning with *and* are described as **adjunctive general extenders** and basically signal that "there is more" (that could be said) and those beginning with *or* are **disjunctive general extenders** that signal that "there are other possibilities" (that could be mentioned)." (**OverstreetOverstreet2021**)

In this example, an oblique phrase is formed with the general noun *baʔe* followed by an adposition and a locative marker, and a second oblique phrase is formed with a specific noun, *bokaɾ*, also followed by an adposition and a locative marker. The general noun and the specific noun are clearly coreferential.

These occurrences of *baʔe* correspond to the definition of placeholders by **Fox2010**, which states that "they fulfill the syntactic projection of the turn so far, rather than simply delaying the next word due, in many cases carrying appropriate nominal or verbal morphology."

An examination of the form of the eight cases at hand shows that the constituent involving *baʔe* and the delayed constituent need not be contiguous. Sometimes the verb is repeated along with the nominal/postpositional phrase, at other times the delayed constituent is placed after the verb as an afterthought, as in (41).

(41) *kiɾ (1.3) tou **baʔe-poɾi=ne**   o-ho o-ʔa, (0.5) kaʔiboʔi-poɾi*
      IDEO   IDEO **thing-at=CONTR** 3-go 3-fall   spirit-at
      'It falls **next to the whatchamacallit**, next to the spirit.'   [22.072]

In terms of morphology found on the placeholders, **Podlesskaya2010** claims that the degree of mirroring of the grammatical marking of the delayed constituent on the placeholder varies depending on the language: it can be total, absent, or partial. All of the Teko cases, including (40), show total mirroring: the placeholder and the delayed noun show equivalent morphology, in a mirror construction. Actually, in two of these examples, the phrase with *baʔe* contains even more material than the phrase with the specific noun, a situation not accounted for in Podlesskaya's typology of mirroring. In both cases, the additional morphology involves a second-position clitic =*ne*, which expresses a contrast with the expectations (**Rose2011**).[18] As a second-position clitic (**Rose2011**), it cannot structurally occur in any other position, and hence cannot be recycled on the specific noun when it is post-verbal, as in (41).

In its placeholder use, I analyze *baʔe* as a noun because

- It is the head of a noun phrase – in (40) and (41) an object of a postposition;

- It is referential, with its usual meaning: its referent is non-human, either inanimate, as in (40), or non-human animate (a non-human being, as in (41) or an animal in other cases).

---

[18]In one of these two cases, the intensive particle =*eʔe* is additionally found on the placeholder and not replicated on the delayed constituent.

Seraku2022 identifies three functions of placeholders. A speaker will use a placeholder "when (i) she has no immediate access to the target form but the time-linear nature of communication forces her to utter something, (ii) she, though aware of the target form, prefers not to reveal it for social reasons, or (iii) she, though aware of the target form, prefers to reveal it at a later stage for rhetorical reasons." The use of placeholders in the Teko corpus fall under the first and the third reasons, i.e. formulation trouble and rhetoric.

Four of the eight examples can be associated with formulation trouble. First, common cues of disfluencies are found in these examples: false starts, repetitions, and a long pause before the target. This is precisely the case of (40), which includes several pauses. It is rendered in Figure 4, with a pause preceding the placeholder, another one following it and preceding the target, and no lengthening of *baʔe*. Unfortunately, given the small number of examples, a statistical analysis comparing prosodic cues (lengthening, pause duration) with the characteristics of both *baʔe* nouns and *baʔe* hesitators is not reliable. Second, it can be noted that the referent of the delayed constituent is always new in discourse, i.e. it may require word search. And more specifically, among the three delayed constituents (one is preceded twice by a placeholder), one is a proper name, a "prototypical search object" (**Schegloff1979**), and the other is a loanword, a type of constituent that a native speaker getting recorded by a linguist might try to avoid by finding a native word.[19] Third, the placeholder use has not been found in the written part of the corpus, where writers had more time to think about formulation (**Seraku2022**).

The other four examples are quite different, in that they show no signs of disfluency. However, all the referents of the delayed constituents are also new in discourse, namely in the one creation myth in which all four examples are found. In this story, people high up in a tree throw down seeds, and when each of these touches the ground it turns into a different animal. In one case, a seed falls down next to a spirit, see (41). It is very plausible that in these four cases, the placeholder is used to create suspense (**Luelsdorff1995**) by postponing a bit the naming of the living entity under focus.

The question remains open whether *baʔe* is sometimes used as a placeholder without the target being realized. There is at least one case in the corpus that could be interpreted as a placeholder without an overt target. Example (42) is an excerpt from a chaotic conversation, where a speaker asks her mother to tell

---

[19]In that specific example, formulation trouble is linked to the second function of placeholders, i.e. not revealing a target form for social reasons (here, expectations of the hearer in terms of "nativeness"). In this example, the speaker failed to find a native term for the glass jar.
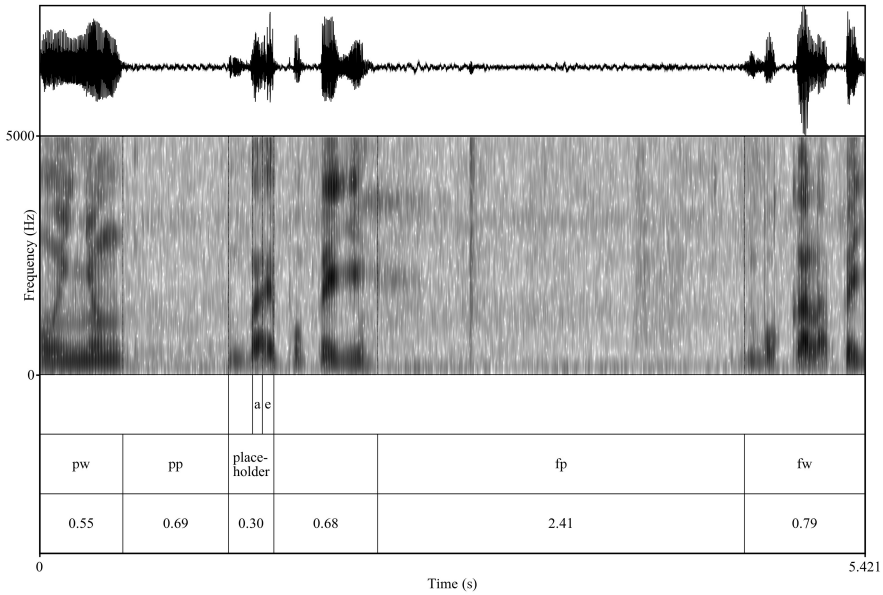
Figure 4: Praat analysis of an occurrence of *baʔe* as a placeholder

me a story. Her request was not followed, so it is unclear what character was targeted. Still, *baʔe* here is possibly a placeholder, uttered to avoid the name of a dispreferred (maybe threatening) entity.

(42) *aʔe* **baʔe** *d-o-aɨhɨ-ʤi-maʔẽ-pe*   *ɛɾe.*
      DEM **thing** NEG-3-like-NEG=REL-about 2SG.say
      'Tell (the story) about the **one** we don't like.'   [31.024]

When *baʔe* is not followed by a target, it might not always be possible to make a straightforward distinction between the uses of *baʔe* where the speaker signals that they are not able or willing to provide a target form and those where it has a clear specific referent, as in (7) to (9).[20] Additionally, it has been shown that sometimes *baʔe* does not have a specific referent, but its use is nevertheless not to signal that the referent should be retrieved for the message to be fully conveyed, as in (12) to (14).

---

[20]For a similar difficulty with *thing* in English, see **Pertejo2015**.

4.3 Rhetorical *baʔe* + N phrase

A different pattern involving *baʔe* is a sequence of *baʔe* and a specific noun, importantly within a single noun phrase and without any pause between them, pointing to a single referent. There are four examples of this pattern in the audio corpus, and two more in the written corpus. Example (43) is one of them.

(43)   o-ʔaɾ-a=itʃe   koɾ   (0.2) **baʔe** ʤaiwəɾ-a-ɾ-aʔɨɾ
       3-be_born-REF=IRR then   **thing** demon-REF-RELN-son
       'It is born. It is then the (thing) son of a demon.'   [22.015]

The examples are few, but striking. In this pattern, the use of *baʔe* is not morphosyntactically required, as it does not fill an obligatory syntactic slot. Yet it might host morphology (there is one example where *baʔe* carries the plural marker). Moreover, semantically, the term *baʔe* does not add any information, as it is a hypernym for the specific nouns that follow it. There are furthermore no cues for disfluency in the audio examples: no striking pause, false start or lengthening of *baʔe*. Figure 5 shows how *baʔe* in (43) is immediately followed by the next noun.

These facts speak against analyses as hesitator, placeholder or N-N compound. The analysis as hesitator can be dismissed when there is morphology on *baʔe* and in the absence of disfluency cues. The analysis as placeholder can be dismissed because a placeholder is expected to "take the place" of the specific noun in the phrase, not to form a complex noun phrase with it. Finally, the analysis in terms of N-N compound can be dismissed on the basis that *baʔe* is a separate word that can take its own morphology, and that the noun following it is not an obligatorily possessed noun as is the case in the N-N compounds (see Section 2.1).

Formally, this sequence of two nouns could be seen as a regular type of noun phrase, made up of a sequence of two nouns, the first one being the head and the second one a modifier (**Rose2011**). Such sequences, exemplified in (44), are morphologically unmarked, form a single prosodic phrase, in which the second noun determines a semantic subset of the possible referents of the first noun (often by gender or age).

(44)   a. *pitaŋ awakʷəɾ*
           child man
           'the boy'
       b. *teko ãbɨɾ*
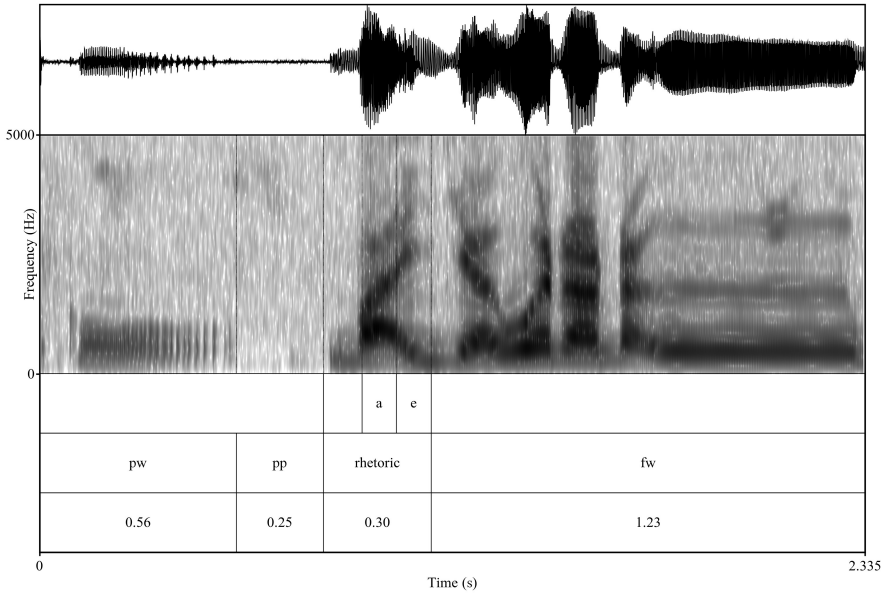   Teko deceased_person
           'the late Tekos'

Figure 5: Praat analysis of an occurrence of the rhetorical use of *baʔe+N*

The difference is that in regular $N_{head}+N_{modifier}$ sequences, both nouns contribute semantically to the identification of the referents, while in all but one *baʔe+N* sequences, see 45, *baʔe* is not informative because the referents of the specific nouns are unquestionably 'non-human'. This raises the question whether this use of *baʔe* might be rhetorical. The fact that it is not semantically informative makes it instrumental in delaying the occurrence of the specific noun, even when speakers do not have a problem in formulating their sentences. Given the small number of examples, two hypotheses can be put forward, though they cannot be verified: *baʔe* might create suspense, or it might be used as a euphemistic or mitigating device.[21] These two hypotheses will be discussed shortly. Interestingly, the two functions of suspense and euphemism are known for being potentially related to fillers, as will be discussed later.[22]

---

[21]A definition of "euphemism" is the following: "A euphemism is used as an alternative to a dispreferred expression" (**AllanAllan1991**).

[22]Brigitte Pakendorf pointed out a description of the Tungusic language Sibe, in which the demonstrative-derived fillers have functions related to both suspense and euphemism, namely "to slow down the speech to introduce [an] important word" or "to express the speaker's reluctance before uttering [a] rather disturbing word" (**Zikmundová2013**).

Regarding the hypothesis that the use of *baʔe* creates suspense, like pauses (and filled pauses) might do, it can be related to the conspiratorial use of fillers "either to prevent potentially overhearing third parties from understanding, and/or to create a collusive air between interlocutors " (**Enfield2003**), which may create suspense. The hypothesis fits well with the fact that in five of the six cases under study, the referent is new: it has not been mentioned in the preceding stretch of text. It also fits well with the fact that these cases occur in sentences expressing surprising information, as is the case when revealing the identity of a participant in (43). This dramatic sense is strengthened by prosody in particular in (43): Figure 5 shows that the word preceding *baʔe*, the conjunction *koɾ* 'then' annotated "pw" in Figure 5 is very much lengthened, delaying the utterance of the noun phrase, and that the last syllable of the noun phrase *ʤaiwəɾ-a-ɾ-aʔiɾ* is much lengthened, putting a very clear emphasis on this term.

As mentioned at the beginning of this paper (Section 2.1), there is one example where *baʔe* precedes a noun phrase for a human being, given in (10) and repeated here in (45).

(45) *si-ɾo-nan   aŋ **baʔe** panaisĩ   wãĩwĩ-am*
      1INCL-SOC.CAUS-run DEM **thing** White_person woman-TRANSF
      'We have kidnapped this white girl.'   [10.030na]

The audio for this text has long been lost, and it is now difficult to justify the original presence of a colon in the transcription between *baʔe* and *panaisĩ*: did I add it to account for a prosodic break or for the (supposed) morphosyntactic organization of the sentence with 'white woman' being an apposition to 'this thing'?[23] In any case, the demonstrative here is cataphoric, lending itself well to supporting an analysis of *baʔe* as a device to add suspense. It is therefore tentatively analyzed as showing the rhetorical *baʔe* + N pattern.

As for the hypothesis that *baʔe* might mitigate a face-threatening expression by preposing a general term to delay the dispreferred specific noun, it has already been recognized that fillers might have a euphemistic function, possibly linked to taboo avoidance (**Enfield2003**; **Cheung2015**; **Pertejo2015**). Indeed, in the languages of the world, general nouns like 'thing' are often used euphemistically. As an example, *faire la chose* 'do the thing' in French is a prudish way to refer to sexual intercourse (**Kleiber1987**). It is well known that "general-for-specific" is a common metonymic strategy for taboo items (**Burridge2006**).

Lexical substitution or metaphorical periphrasis is often employed in languages spoken in the Amazon basin, due to taboo avoidance (**DienstDienst2009**;

---

[23]In those days, no Teko speaker knew how to write their language.

Mihas2019; **Wojtylak2015**; **Wojtylak2019**). This is in line with the Amazonian cosmology (**Castro1998**; **Santos-Granero2007**), in which each type of entity (humans, animals, plants, dead people, other beings) must be clearly identified to avoid any confusion between social entities. And indeed, in the six Teko occurrences of this pattern, the referent of the specific noun is always a non-Teko animate, i.e. a socially sensitive entity (from the human point of view). It is either an animal with special features (being able to speak, for example) or another type of human or non-human being. If we are dealing with a euphemistic use of *baʔe* before nouns for animates in Teko, it would technically not be taboo avoidance through lexical substitution but rather taboo delaying by addition of a general term, mitigating the face-threatening power of the name of a potentially dangerous being by starting to refer to it vaguely. In a very similar construction, *poi* in Japanese can be used before genital-denoting nouns to mitigate their face-threatening effect (**Seraku2024**).

What remains to be determined is whether this use results in a "constructional change" (**TraugottTrousdale2013**), i.e. on the basis of an existing noun phrase structure with two nouns, potentially a semantic bleaching of *baʔe* (with (45) analyzed as showing the same pattern), and semanticization of a pragmatic implicature that could read as follows "name of potentially dangerous referent is coming".[24] Given the present state of the corpus and the lack of regular access to consultants, this issue remains to be solved in the future.

5 Diachrony

This section addresses the diversity of *baʔe* elements and their use in Teko from a diachronic perspective. It starts in Section 5.1 by putting forward the nominal source of the hesitator within a rich network of grammaticalizations issued from the noun 'thing'. In Section 5.2, it concentrates upon that specific path of evolution from noun to hesitator, involving the placeholder function as a likely bridging construction. It then presents in Section 5.3 an original hypothesis that it is through the filler use that the Teko verb 'do/make' arose. A summary of the diachronic hypotheses will be presented in the final summary of the paper.

5.1 A general noun as a source for the hesitator

Cross-linguistically frequent sources for fillers are listed in the literature (see for example **Podlesskaya2010**; **Enfield2003**). They are briefly presented below:

---

[24]In fact, the concept of 'spirit' is found as a secondary or a new meaning for reflexes of Proto-Tupí-Guaraní *maʔe in some Tupí-Guaraní languages (**chousouNodate**)

- pronouns: third person, demonstrative, interrogative, indefinite or emphatic, such as Japanese *ano* (**HayashiYoon2006**).

- clauses, often with an interrogative word and naming nouns or verbs, such as English *whatchamacallit* (**Enfield2003**).

- a general noun; examples from **Podlesskaya2010** are Armenian *ban* 'thing', Turkish *sey* 'thing' and Vietnamese universal classifier for objects *cai*.

- both a demonstrative and a general noun, as found in Nahavaq *taqtak* 'this thing' (**Dimock2010**) and Lao *qan⁰-nan⁴* 'that thing' (**Enfield2003**).

- both a nominal placeholder and a general verb, like the Udi combination *he-b-* made of the non-human nominal placeholder *he* and the verb *b-* 'do' (**GanenkovGanenkov2010**).

General (or generic) nouns are known as good sources for grammaticalization. In their *World Lexicon of Grammaticalization*, **KutevaKuteva2019** list five grammaticalization targets of the noun 'thing': complementizer, indefinite pronoun, attributive possession marker, nominalizer, and interrogative pronoun. And indeed, **AuweraAuwera2021** have pointed to a number of grammaticalization paths from the noun *maʔe 'thing' in Tupí-Guaraní languages. The general picture of that network of changes is schematized in Figure 6.[25]

In this figure, discourse markers refer to Nheengatu clause-final 'protest' *baʔ*, Siriono *ba* surprise particle, Warázu *maʔe* 'yes', Mbyá Guaraní *mbaʔe* particle used to either introduce a proposal with a 'how about' meaning, or render 'for example, perhaps' (**AuweraAuwera2021**: 77–78). The authors posit that these various "discourse markers" arose diachronically from the "interrogative" marker, rather than from its original nominal source.

The same question holds for the Teko hesitator: was it derived from the noun 'thing', or from one of the other Teko *baʔe* forms described in Section 2 (nominalizer, interrogative, or verb), for which diachronic evidence suggests that they themselves derive from the noun? Basically, the question is whether the noun 'thing' is the direct source of the hesitator, or its ultimate source, with another item constituting an intermediary step in the diachronic development. Let us consider these different routes.

---

[25]The figure distinguishes "between constructions that are crucially non-human (unmarked), constructions that may be human or non-human (italics) or constructions in which the feature has become irrelevant (in boxes)" (**AuweraAuwera2021**). The question mark after "nominalizer" is present in the original figure and represents remaining issues in the diachronic analysis (**AuweraAuwera2021**).
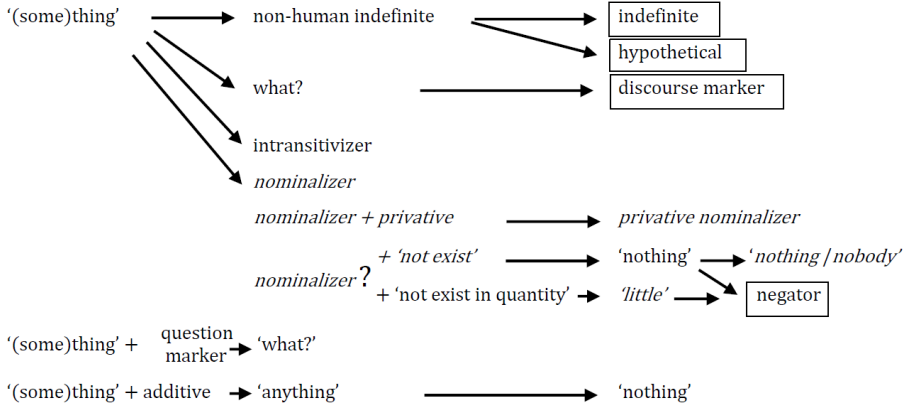
Figure 6: Diachronic developments from the noun 'thing' in Tupí-Guaraní languages (copied from **AuweraAuwera2021**)

First, there is no morphological, syntactic, or semantic reason to suspect the nominalizing use of *baʔe* to have given rise to the hesitator.

Second, the possibility that the verb 'do/make' would have given rise to the hesitator is not theoretically or typologically implausible. Indeed, its general meaning could facilitate its use as a filler, and 'do' verbs have been recruited as fillers in some languages (see above). However, the fact that only two Tupí-Guaraní languages have a verb formally related to the noun 'thing', while a greater number seem to display the filler use is not very favorable to this path of evolution (at least not in that direction, see Section 5.3).

Third, the strongest hypothesis for an intermediary step in the evolution from the noun to the hesitator would be the interrogative, similarly to what **AuweraAuwera2021** posit for discourse markers in the Tupí-Guaraní language group. A caveat is that for now there is no robust argument to ascertain that there is an interrogative element *baʔe* distinct from the noun in Teko. And there is no evidence for interrogative as an intermediary stage in the development of the hesitator: there is no morphological, prosodic or pragmatic trace of interrogation in the use of the hesitator.

Consequently, it seems more plausible at this point to trace back the hesitator to the noun *baʔe*. Needless to say, there is no doubt about the direction of evolution, given the reconstruction of the noun at the group level.

In that perspective, the following diachronic mechanisms must have been at work in the reanalysis of the original noun into a hesitator:

- semantic bleaching: loss of the 'nonhuman' semantic feature;

- reanalysis: loss of nominal morphology;

- extension: wider distribution (not just within noun phrase);

- gain of specific prosody (lengthening, pauses).

5.2 From noun to hesitator via placeholder

One can suppose that the discourse uses of *baʔe* have built bridges between its basic nominal use and its status as a hesitator.

A first diachronic step would have involved the development of the discourse uses of the general noun. Although **Podlesskaya2010** asserts that placeholders may have uses as vague or generic expressions and vague category identifiers (i.e. our general extenders), in the case of Teko there is clear comparative evidence that the original element was a general noun. Remember that its actual meaning as a noun is 'non-human' rather than restrictively 'thing', covering inanimates, animals and other non-human beings. The wide range of meanings that this general noun covers has very likely facilitated its use as a vague expression, in either its extender, nominal placeholder, or rhetorical use. The extender and placeholder uses retain the original non-human meaning of the general noun *baʔe* but systematize its potential as a vague expression, which is not systematically implemented in its basic nominal use (see Section 2.1). As **HenneckeHennecke2022** put it for French placeholders: "This particular placeholder function arguably arises in a process of subjectification encoding speaker attitudes and intersubjectification, in this case the request for completion by the listener."

In a subsequent diachronic step, the vagueness component present in the discourse uses of the general noun becomes semanticized in the hesitator. In particular, the placeholder use with its intrinsic expectation for further specification opened the way to the "cataphoric" use of *baʔe*, a use where the vagueness expressed by the general noun is expected to be solved by a further specification of the referent. This expectation that something more is to follow, a plain inference for the general noun, ends up being part of the core meaning of the hesitator, in a process sometimes referred to as pragmaticalization. Pragmaticalization has indeed been defined as "the process by which a syntagma or word form, in a given context, changes its propositional meaning in favor of an essentially metacommunicative, discourse interactional meaning" (**Frank-Job2006**). It is often viewed

as a process similar to grammaticalization, the endpoint of which is not a "grammatical" element but a pragmatic or discourse marker (see **Diewald2011**).[26]

HenneckeHennecke2022 concretely consider for French hesitators that "the use of syntactically integrated placeholders immediately followed by the sought for target may be easily reanalysed as fillers [hesitators in the terminology of this volume] interrupting the syntactic construction in order to gain time, possibly the bridging context between placeholders and fillers [hesitators]." This scenario is very plausible for Teko as well. The development of hesitators via placeholder uses has also been proposed for Mashti (**chapters/rice**) and Amazonian Spanish (**VallejosYopán2023**).

In the process of pragmaticalization, the types of target of *baʔe* widened from non-human nouns to constituents of all sorts and length (see **Mihatsch2006**: 166 on a parallel process in French), thus losing its morphosyntactic characteristics as a noun and its non-human semantic component. Consequently, it lost all its nominal characteristics and developed into a new part of speech.

5.3 Further development as a verb

In Section 2.4, it was said that *baʔe* as a verb is used to mean 'make', 'create', 'act', and as a demonstrative manner verb: it thus covers a range of specific and general meanings. It is tempting to explain this synchronic variation by a semantic shift of bleaching from specific to general meanings. However, it was also shown that this verb was not inherited from Proto-Tupí-Guaraní but is an innovation in Teko (as well as Kaapor), while the noun for 'thing' has been reconstructed as **maʔe* for Proto-Tupí-Guaraní. Consequently, it seems worth investigating whether the verb *baʔe* could have emerged from one of the filler uses: the placeholder or the hesitator.

Actually, several European languages show the derivation of a nominal placeholder related to the noun for 'thing' into a verbal placeholder: Brazilian Portuguese *coisa* into *coisar* and Italian *coso* (cf. *cosa* 'thing') into *cosare*, and Dutch

---

[26]This paper will not address whether the shift from noun to hesitator should be considered a case of grammaticalization. First of all, the literature on fillers is not consistent on their status as grammatical elements: they are described as discourse markers (**Podlesskaya2010**), as not being discourse markers (**VallejosYopán2023**), or as grammatical items (**Kirjavainen2022**), respectively. Second, the literature on pragmaticalization is not consistent in its relationship to grammaticalization (**Diewald2011**; **Heine2013**; **DegandDegand2015**). "It is shown that the different positions encountered in the literature can be brought back to diverging views on the conceptualization of grammar, the categorization of discourse markers, and the weight that is put on specific processes involved in the diachronic change." (**DegandDegand2015**). This discussion therefore goes far beyond the descriptive scope of the present paper.

dinges (from *ding* 'thing') into *dingesen* (Wikipedia contributors 2023). Examples (46) and (47) show the nominal and verbal placeholders of Brazilian Portuguese (courtesy of Emerson José Silveira da Costa on the Etnolinguistica list —a South Americanist linguistics mailing list— on March 17th 2022). This is also the case with the Nahavaq placeholder mentioned in Section 5.1, based on the noun 'thing' and a demonstrative, which serves as both a nominal and a verbal placeholder (**Dimock2010**).

(46) *Dá-me   aquel-a **coisa**.*
      hand-me DEM-F   **thing**(F)
      'Hand that thing over to me.'

(47) *Você **cois-ou**   o-s   documento-s?*
      2SG **PH**-2/3SG.PAST the.M-PL document(M)-PL
      'Did you thingummy [staple] the documents?'

This raises the question whether the verb *baʔe* in Teko could have originated from the placeholder uses of the noun for 'thing'. A major problem for this hypothesis is that, contrary to European languages, conversion from nouns to verbs (or vice versa) is not a productive pattern in Tupi-Guarani languages. In fact, the Teko lexicon does not show any verbo-nominal term, undermining the conversion hypothesis.

Another line of investigation is that the verb *baʔe* could have developed from the hesitator, in Teko as well as in Kaapor, the second Tupí-Guaraní language in which a reflex of *maʔe is said to be used as the verb 'do', and remarkably a language for which the use of *maʔɛ* as a hesitator has also been noticed (**Godoy2020**). Theoretically, the use of the hesitator before verbs (creating expectations in a cataphoric way) could have developed into the manner demonstrative verb and from there on to the general meaning 'do, act'. There are nevertheless no good examples serving as a bridging context, because in the use most closely associated with the manner demonstrative meaning, that of introducing ideophones, the verb *baʔe* comes after the ideophones, rather than before (see (25)). The development from hesitator to verb thus remains speculative at this point.

In any case, once used as a verb, *baʔe* would have gained the possibility to combine with verbal morphology, in parallel with the nominal morphology found in its nominal counterpart (including the placeholder use). The other meanings of the verb would have then arisen by gaining more specific meanings. First, a general verb meaning 'act' could arise by extending the use of the manner demonstrative verb to non-deictic use, i.e. contexts/co-text where the action is not visible/mentioned. One would also need to postulate later semantic enrichment, by

which it gained the semantic component of 'creation' and 'manipulation' to extend into the 'create/make' meanings.

6 Summary and conclusion

This paper has offered a new description of the discourse functions of the *baʔe* form in Teko, covering both discourse uses of the noun 'thing' and the distinct hesitator word. It has in particular insisted on the semantic, morphosyntactic and prosodic features of 103 *baʔe* tokens in a corpus of audio-recorded texts: the investigation clearly shows that the hesitator is a distinct part of speech from the noun. It lost its non-human meaning and nominal morphosyntax and gained prosodic features typical of disfluencies.

The analysis was presented on the background of a rich network of evolution from the Proto-Tupí-Guaraní 'thing' noun within the language group, and in Teko in particular. This led us to draw a tentative scenario of evolutions from the noun 'thing' to other homophonous entities in Teko. The diachronic hypotheses put forward in Section 5 are summarized in Figure 7.[27]

In this paper, three developments of the noun 'thing' in Teko were put forward. First, I have described how the noun 'thing' takes on the functions of extender, placeholder, rhetorical use, and interrogative in discourse. Second, I have argued that the noun for 'thing' has developed into an independent hesitator, very likely via the placeholder use of *baʔe*. Third, it has been suggested that the verb 'do, make' may in turn have derived from the hesitator.

As mentioned in Section 1.3, this study was conceived as a exploratory study of fillers in Tupí-Guaraní languages, with the hope that it will lead the way to similar studies in Tupí-Guaraní languages other than Teko.

Acknowledgements

---

[27]Constructions used with non-human referents are unmarked, those with either human or non-human referents are italicized, and those for which humanness is irrelevant are presented in boxes. Dotted arrows represent special contextual uses of the item preceding the arrow, without a change in part of speech.
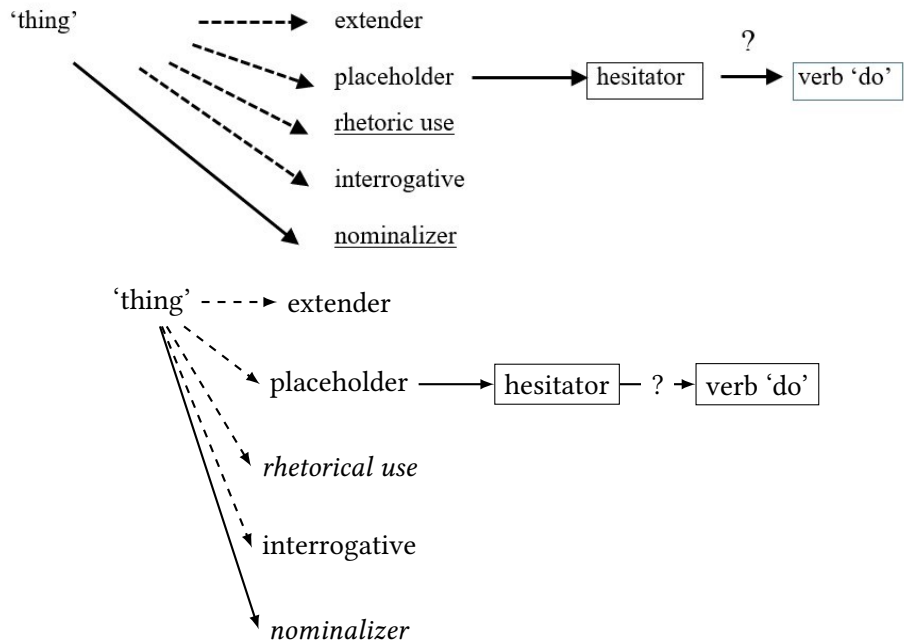
Figure 7: Diachronic developments from the noun 'thing' in Teko

Abbreviations

COMPL	completive	INJ	injunctive
CONCES	concessive	INTER	interrogative
CONT	continuative	INTENS	intensive
CONTR	contrast	IRR	irrealis
DEM	demonstrative	LOC	locative
DIM	diminutive	MAN.DEM	manner demonstrative
EXCL	exclamative	NEG	negation
EXH	exhortative	NSP	non-specific possessor
EXIST	existential copula	PAST	past
FUT	future	PH	placeholder
HES	hesitator	PL	plural
IDEO	ideophone	PL.S	plural of subject
IMP	imperative	PL.O	plural of object
INCL	inclusive	POSTP	postposition
INDET	indeterminate	PFV	perfective

PURP	purpose	SG	singular
RED	reduplication	SOC.CAUS	sociative causative
REF	referential	SUB	subordinator
REL	relativizer	TRANSF	transfer
RELN	relational		