

Fillers

Hesitatives and placeholders

Edited by

Brigitte Pakendorf

Françoise Rose

Research on Comparative Grammar 5



Research on Comparative Grammar

Series editor

Martin Haspelmath (Max Planck Institute for Evolutionary Anthropology)

In this series:

1. Bahrt, Nicklas N. Voice syncretism.
2. Veselinova, Ljuba & Arja Hamari (eds.). The Negative Existential Cycle.
3. Janic, Katarzyna, Nicoletta Puddu & Martin Haspelmath (eds.). Reflexive constructions in the world's languages.
4. Persohn, Bastian. The polyfunctionality of 'still' expressions: A cross-linguistic exploration.
5. Brigitte Pakendorf & Françoise Rose (eds.). Fillers: Hesitatives and placeholders.

ISSN (print): 2749-7801

ISSN (electronic): 2749-781X

Fillers

Hesitatives and placeholders

Edited by

Brigitte Pakendorf

Françoise Rose



Brigitte Pakendorf & Françoise Rose (eds.). 2025. *Fillers: Hesitatives and placeholders* (Research on Comparative Grammar 5). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/484>

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: 978-3-96110-526-7 (Digital)

978-3-98554-146-1 (Hardcover)

ISSN (print): 2749-7801

ISSN (electronic): 2749-781X

DOI: 10.5281/zenodo.15632051

Source code available from www.github.com/langsci/484

Errata: paperhive.org/documents/remote?type=langsci&id=484

Cover and concept of design: Ulrike Harbort

Typesetting: Sebastian Nordhoff

Proofreading: Carmen Jany, David Carrasco Coquillat, Elisa Roma, Eva Schultze

Berndt, Jeroen van de Weijer, Killian Kiuttu, Lachlan Mackenzie, Maria Onoeva,

Nicoletta Romeo, Jean Nitzke, Rainer Schulze, Mary Ann Walter, Yvonne Treis

Fonts: Libertinus, Arimo, DejaVu Sans Mono, Source Han Serif JA

Typesetting software: X_ELA_TE_X

Language Science Press

Scharnweberstraße 10

10247 Berlin, Germany

<http://langsci-press.org>

support@langsci-press.org

Storage and cataloguing done by FU Berlin



Contents

1	Fillers in the world's languages: A refined typology Brigitte Pakendorf & Françoise Rose	1
2	Placeholders versus general extenders in Tungusic languages Elena Klyachko	41
3	A corpus-based investigation of fillers in Negidal (Northern Tungusic) Brigitte Pakendorf	75
4	Copula or placeholder: A qualitative and quantitative analysis of <i>xe-</i> in Kolyma Yukaghir discourse Albert Ventayol-Boada	131
5	Choosing fillers in Besemah Bradley McDonnell & Blaine Billings	169
6	Fillers in the gray: Indeterminacy in their use in everyday conversation Blaine Billings & Bradley McDonnell	211
7	Form and frequency of Kalamang fillers Eline Visser	251
8	<i>That</i> placeholder in Komnzo Christian Döhler	275
9	Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan) Maïa Ponsonnet	315
10	Fillers and beyond Marianne Mithun	341

Contents

11	<i>Mashti</i>: A multipurpose filler in Northern Pastaza Kichwa	
	Alexander Rice	373
12	One more thing ‘thing’ can do in Tupí-Guaraní languages: The Teko filler	
	Françoise Rose	441
Index		485

Chapter 1

Fillers in the world's languages: A refined typology

Brigitte Pakendorf^a & Françoise Rose^a

^aLaboratoire Dynamique du Langage (CNRS & Université Lumière Lyon 2)

Interest in placeholders and related items in the languages of the world has burgeoned in recent years, both from a descriptive and a theoretical perspective. A particular focus of the recent literature has been on the pragmatic extensions of placeholders, demonstrating that these are not merely used in situations of disfluency to substitute for words that momentarily elude the speaker, but are frequently used intentionally to avoid terms for reasons of politeness, “conspirational” motivations, or rhetorical purposes.

Fillers – a term that subsumes both placeholders and hesitatives – are cross-linguistically widespread, even though dedicated studies of such items are notably lacking for the languages of Africa. The distinction between placeholders and hesitatives is one of referentiality and morphosyntactic integration: placeholders are both referential and morphosyntactically integrated while hesitatives are neither. We distinguish five different types in our extended typology: 1) specific placeholders that substitute for a particular part of speech, 2) general placeholders that are not restricted to a particular part of speech, 3) general hesitatives such as the cross-linguistically common “pause vowels”, 4) specific hesitatives found in some Austronesian languages that are not referential or integrated, but are specific to particular delayed constituents, and 5) versatile fillers that can fulfill both placeholder and hesitative functions.

The in-depth studies of individual languages collected in this volume highlight the fact that it is not always straightforward to distinguish particular instances of a filler from those of its source form nor to distinguish hesitative from placeholder uses of versatile fillers, that placeholders are frequently not followed by an overt target, and that filler use can vary greatly among speakers.

keywords: placeholders, hesitatives, filled pauses, disfluency, watchamacallit, avoidance



Brigitte Pakendorf & Françoise Rose. 2025. Fillers in the world's languages: A refined typology. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 1–39. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697575

There are marginal phenomena that give a basis for very important lessons about the nature of language in general, about particular languages, and about the way we need to approach the examination and analysis of languages.

(Joseph 1997: 200)

1 Introduction

This volume focuses on a marginal phenomenon that is starting to garner more and more attention, namely fillers.¹ These are markers of hesitation like *uh/uhm* and substitutes for items that elude a speaker or that she does not want to utter, like *watchamacallit* or *thingy*; as such, they belong to the many vague expressions found in natural languages (Cutting 2007). In this volume we focus on conventionalized fillers² that roughly coincide with what Bloomfield (1933: 186, cited from Maclay & Osgood 1959: 21) called “special parenthetic hesitation forms”. *Ad hoc* delaying devices, such as silent pauses, final syllable lengthening, or false starts and repetitions, as well as discourse particles that have a wide range of functions, are thus excluded from consideration (other than in comparison with conventionalized fillers). We define fillers as follows:

- (1) A FILLER is an overt and independent marker of hesitation with a conventionalized form which is used in disfluency and which often has other discourse functions. Two strategies of use can be distinguished, namely HESITATIVES (or HESITATORS), elements that are non-referential and not morphosyntactically integrated into the unfolding utterance, and PLACE-HOLDERS, which substitute for a particular element of the utterance and hence are morphosyntactically integrated and referential.

For instance, French *eum* is a hesitative: it is not referential and not morphosyntactically integrated (1). In contrast, Dalabon *keninjhbi* is a placeholder, since it is referential and integrated within a noun phrase (2).

¹The term ‘filler’ is used in different ways in the literature: it was introduced by Clark & Fox Tree (2002: 75) with reference to English *uh* and *uhm*, which had until then been mostly designated as ‘filled pauses’ (a label that continues to be used for items like *uh* and *uhm*, cf. Lickley 2015: 458, Kosmala & Crible 2022, Kirjavainen et al. 2022, in addition to ‘filler’). Sometimes, however, the term ‘filler’ is used much more broadly as a label “... for all of the non-silence devices that can be deployed after the current word has been brought to completion to delay the next word due” (Fox 2010: 2), i.e. it includes final vowel lengthening or discourse particles.

²Note that some “pause vowels”, such as *uh* in English or *eh* in Spanish are conventionalized (cf. Clark & Fox Tree 2002, Erker & Vidal-Covas 2022).

- (1) French (Kosmala & Crible 2022: 7)

une anecdote t' en as pas une? eum (1.120) *ah,*
 an anecdote 2SG PART.O have[PRS.2SG] NEG one **uhm** ah,
c'est bon!

that.is good

'an anecdote don't you have one? **uhm** (1.120) ah I have one!'³

- (2) Dalabon (Ponsonnet 2025 [this volume]: 322, ex.(9))

Kanidjah bala-lng-bo-ninj keninjhbi-ngong [3.76s] *kunj-ngong*
 DEM 3pl-SEQ-go-PIMP PH-group kangaroo-group
bala-h-bo-ninj keninjhbi djakana...
 3pl-R-go-PIMP PH bird.species

'They were going there, all the whatsit... [pause] all the kangaroos were going there, the whatsit jacanas.'

Such fillers were at the margins of linguistic research for a considerable time, with notable exceptions being the seminal paper by Hayashi & Yoon (2006) and the edited volume by Amiridze et al. (2010). Recently, however, interest has exploded, with a spate of papers exploring fillers in different languages appearing in the last few years (e.g. Seraku 2020, Hennecke & Mihatsch 2022, Klyachko 2022, Nagaya 2022, Vallejos Yopán 2023). Seraku, in particular, has explored the pragmatics of placeholders in the light of different semantic and formal approaches (Seraku 2022a, 2022b, 2023, 2024). However, the preliminary typology proposed by Podlesskaya (2010) still remains the main cross-linguistic overview of morphosyntactic aspects of fillers (more precisely, of placeholders), and we believe the time has come to bring our knowledge on all aspects of fillers up to date. Hence we aim to provide a review of both functional and morphosyntactic aspects of fillers in a cross-linguistic perspective, basing ourselves on published literature as well as on the new insights that emerge from the chapters included in this volume. As a result we propose a refined morphosyntactic typology of fillers (Table 1, see §4 for details).

The rest of the article is structured as follows: §2 provides an overview of what is known about the cross-linguistic distribution of fillers, and §3 discusses the diachronic sources of these elements. In §4 our refined morphosyntactic typology of fillers is outlined, while §5 deals with the diverse functions of these items and §6 with their relation with general extenders. The subsequent sections briefly

³In the examples, fillers are highlighted in bold, while their targets, where relevant, are underlined.

Table 1: Refined typology of fillers

specific placeholders	specific hesitatives
general placeholders	general hesitatives
versatile fillers	

touch upon fillers in language contact (§7) and what they can tell us about the nature of language (§8), and §9 describes the volume and its main contributions to the study of placeholders and hesitatives.

2 Fillers in the languages of the world

Fillers are known from practically every continent. The data ranges from brief mentions in grammars – possibly illustrated with a couple of examples – to in-depth studies of various aspects of their forms and uses.

In fact, fillers are rarely discussed within grammars. A search was realized in a reduced but likely representative sample of grammars (Lahaussois to appear). Within the table of contents of 237 grammars, our query resulted in only nine successful hits: six for “filler”, two for “hesitation”, and one for “placeholder”. The key terms “hesitator”, “hesitative” and “filled pause” were not found at all. Although this search may miss grammars that do discuss fillers but use different labels, it does indicate that fillers are mostly considered a marginal question when describing a language. The hits refer to seven grammars in total (two grammars use two key terms in their table of contents), of which the oldest was published in 1993. This indicates that fillers have only recently emerged as a topic worthy of description in grammars. These seven grammars show a fair level of cross-linguistic coverage, with three languages from the Eurasian macro-area – Modern Welsh (King 2003), Dumi (van Driem 1993), and Japhug (Jacques 2021) – one language from the Papuan macro-area (Papuan Malay; Kluge 2017), one from Australia (Kuuk Thaayorre; Gaby 2017), one from South America (Movima; Haude 2006), and one from Africa (Tiefo-D of Daramandugu; Heath & Ouattara 2021). Only North America is not represented.

Dedicated studies on fillers also cover quite a number of languages, albeit with a bias towards the Eurasian macro-area. Fillers are well described in several East Asian languages: Japanese (Kitano 1999, Hayashi & Yoon 2006, Seraku 2022a, Seraku 2022b), Yoron Ryukyuan (Seraku 2020), Korean (Hayashi & Yoon 2006),

and Mandarin Chinese (Zhao & Jurafsky 2005, Hayashi & Yoon 2006, Cheung 2015). They are also fairly well described for various languages of Northern Asia: Kolyma Yukaghir (Ventayol-Boada 2025 [this volume]), the Tungusic languages Even, Evenki, and Negidal (Matić 2008, Klyachko 2022, Klyachko 2025 [this volume]), Pakendorf 2025 [this volume]), and the Turkic language Dolgan (Däbritz 2018), as well as for some languages of the Caucasus: the Nakh-Daghestanian languages Udi, Agul (Ganenkov et al. 2010), and Rutul (Maisak 2023), as well as Armenian (Khurshudyan & Podlesskaya 2006) and Georgian, which has a placeholder verb (Amiridze 2010). In European languages, they are known from English (e.g. Enfield 2003, Palacios Martínez & Núñez Pertejo 2015, Tárnýiková 2019), which is often the benchmark for comparison with other languages, and have been described in some detail for French (Mihatsch 2006, Hennecke & Mihatsch 2022, Corminboeuf & Johnsen 2023), Russian (Podlesskaya & Kibrik 2009, Podlesskaya & Korotaev 2022), German (Vogel 2020), Danish (Navarretta 2016), and Estonian (Keevallik 2010). As for Spanish, Bajo Pérez (2019) and Mihatsch (2024) describe two different fillers, *fulano/fulana* and *chisme*, respectively, while Vallejos Yopán (2023) analyses the filler *este* in the variety of Spanish spoken in Amazonia.

Other dedicated studies on fillers in South America cover Northern Pastaza Kichwa (Rice 2025 [this volume]) and the Tupí-Guaraní language Teko (Rose 2025 [this volume]). As for North America, fillers have been described in fair detail in a handful of languages: Maliseet-Passamaquoddy (LeSourd 2003), Sliammon Salish (Watanabe 2010), and Mohawk (Mithun 2025 [this volume]). Very little work has been dedicated to fillers in Australian languages; a start has been made by Ponsonnet (2025 [this volume]), who provides a detailed discussion of the Dalabon placeholder *keninhbi*. In contrast, Austronesian fillers have generally been described in more detail, possibly because some of them show quite striking formal features (see §4.4 for details) and because a filler is reconstructed to the proto-language: **anu* ‘whatchamacallit’ (Nagaya 2022: 92, citing Blust 2013)⁴. Thus we have descriptions of fillers for Nahavaq spoken in Vanuatu (Dimock 2010), Ilocano (Rubino 1996), Western Subanon (Blake 2020), and Tagalog (Nagaya 2022) of the Philippines, and for various languages spoken in Indonesia, such as Besemah (McDonnell & Billings 2025 [this volume]) and Nasal (Billings & McDonnell 2025 [this volume]) spoken on Sumatra, as well as Indonesian itself (Wouk 2005).

⁴In the web edition of the Austronesian Comparative Dictionary (Blust & Trussel 2013, https://www.trussel2.com/ACD/acd-s_n1.htm?zoom_highlight=whatchamacallit, accessed 26 July 2024), the authors define the Proto-Malayo-Polynesian form **a-nu* as “thing whose name is unknown, avoided, or cannot be remembered: what?”.

A surprising exception is Africa, for which no detailed descriptions of fillers seem (yet) to exist. Even though absence of evidence is not evidence of absence (all the more so since Heath 1999, 2008 and Heath & Ouattara 2021 briefly describe fillers in three unrelated languages of West Africa), the lack of dedicated descriptions of fillers in African languages is notable and possibly indicates a relative paucity of conventionalized placeholders and hesitatives in this region.

This brief overview demonstrates that – with the exception of some large and well-studied languages like Japanese or Mandarin Chinese – to date little is known about fillers in the languages of the world. The papers included in this volume thus fill in considerable gaps in our knowledge of these items, especially for the Americas, Australia, and the non-Austronesian languages of the Pacific. As mentioned above, the aim of the present chapter is to fill the gap of typological studies on fillers, which are notably rare (Hayashi & Yoon 2006, Podlesskaya 2010, Seraku 2024).

3 The origins of the fillers

Some fillers do not have any other function in the current state of the language; these we call DEDICATED. In some cases, an erstwhile lexical or phrasal source was lost or modified in the language. This is the case for the French placeholder *truc*, the origin of which is a word for theatre machinery (Mihatsch 2006), unsuspected by most speakers. But the filler might also lack a lexical source, such as the Yoron Ryukyuan filler *muna* (Seraku 2020). This concerns especially the “pause vowels” (Fox 2010: 2, Candea et al. 2005) used as hesitatives in many languages. Nevertheless, Clark & Fox Tree (2002: 75) “argue that *uh* and *um* are, indeed, English words. By words, we mean linguistic units that have conventional phonological shapes and meanings and are governed by the rules of syntax and prosody”.

Dedicated fillers have been described in various languages around the world (though with labels different from ‘dedicated’ or ‘fillers’): Yoron Ryukyuan (Seraku 2020), Ilocano (Rubino 1996: 656–661), Nahavaq (Dimock 2010: 130–131), Manambu (Aikhenvald 2008: 571–578), and Evenki and other Tungusic languages (Klyachko 2022, 2025 [this volume]). Furthermore, the proto-Austronesian form **anu* can be assumed to have been a dedicated filler (see footnote 4).

In contrast to dedicated fillers, the source of many fillers is still synchronically present in the language. The filler might be perfectly homonymous with its (lexical or phrasal) source form, or its form might have been modified through a process of pragmaticalization, yet the similarity with its source makes the diachronic relation obvious. An example of perfect homonymy is that of the Even

filler *iak*, which is indistinguishable from the interrogative ‘what’ (Matić 2008). An example of a transparent source is that of the French placeholder *machin*. It is obviously formally similar to *machine* ‘machine’ (it originates in a word for war machinery, Mihatsch 2006), even though it differs from it phonologically (/maʃē/ vs /maʃin/) and morphologically (*machin* is masculine, *machine* is feminine). Similarly, the Amazonian Spanish filler *este* is clearly derived from the masculine singular proximal demonstrative, but has developed a different prosodic form (Vallejos Yopán 2023).

Potential sources for placeholders were listed by Podlesskaya (2010). The most common source for placeholders are pronouns, especially demonstrative and interrogative pronouns.⁵ Demonstratives used as fillers are cross-linguistically common: among others, they are frequently used in Japanese and Korean (Hayashi & Yoon 2006, Seraku 2022a), in Austronesian languages (Wouk 2005, Kluge 2017: 388–389, Billings & McDonnell 2025 [this volume], McDonnell & Billings 2025 [this volume]), in Kuuk Thaayorre from Australia (Gaby 2017: 252–254), and in Russian (Podlesskaya & Kibrik 2009). Interrogative-based fillers, too, are prevalent in the world’s languages, having been described for Mandarin Chinese (Cheung 2015), Austronesian languages (Dimock 2010, Kluge 2017, Nagaya 2022, Billings & McDonnell 2025 [this volume]), Even and Evenki (Matić 2008, Klyachko 2022), Dolgan (Däbritz 2018), and Udi and Agul spoken in the Caucasus (Ganenkov et al. 2010). Another common source is that of a semantically bleached noun, often a general noun, such as the Teko hesitative, which originates in a general noun for non-humans (Rose 2025 [this volume]). In Armenian, too, a noun with a general meaning, *ban*, can be used as a filler (Khurshudyan & Podlesskaya 2006), and in Kalamang the general noun functions as a placeholder with avoidance use (Visser 2025 [this volume]; see §5.1). In contrast, French *machin* and *truc* are examples of fillers whose lexical sources were semantically specific to start with. A third common source of fillers are lexicalized constructions, often involving an interrogative and a naming noun or verb. A typical example is English *whatchamacallit*; further fillers with a phrasal source are found in Russian (Podlesskaya & Korotaev 2022), Dalabon (Ponsonnet 2025 [this volume]), and Northern Pastaza Kichwa (Rice 2025 [this volume]). A possible source that has

⁵Note that Podlesskaya (2010: 12) includes personal and indefinite pronouns in her list of pronominal sources of placeholders. These seem to be extremely rare: the only examples we have come across (thanks to Wiltrud Mihatsch) are French pronouns such as *quelque chose* ‘something’ or *quelqu'un* ‘someone’ which are modified by determiners and which function as avoidance placeholders, as shown by the author’s description: “...the pronoun is the substitute for a noun one cannot or will not provide” and “In this respect, the designative use is not unlike *truc*, *machin*, *bidule* [...], and other *thingummy* and *whatsit*.” (Larrivée 2009: 10, 13).

not been mentioned before is a copula: Kolyma Yukaghir spoken in Siberia has an element *xe-* that functions as a copula on the one hand and as a placeholder on the other. The functions can be distinguished on syntactic and intonational grounds, but the precise historical link between these two synchronically distinct items is unclear. However, Ventayol-Boada (2025 [this volume]) argues in favour of the placeholder developing out of the copula.

Podlesskaya mentions that placeholders can also result from a combination of the potential sources mentioned above. In fact, in several languages it is not straightforward to identify which element functions as the placeholder *per se*, since these consist of multiple items. For instance, the Mandarin Chinese complex phrase *na-ge shenme* consists of the distal demonstrative *na* combined with the neutral classifier *ge* followed by the interrogative *shenme* ‘what’. Hayashi & Yoon (2006: 493–494) include this in their discussion of the placeholder use of demonstratives, saying that the distal demonstrative used as a placeholder is frequently followed by the interrogative. In contrast, Cheung (2015) includes the same expression in his discussion of “wh-placeholders” in Mandarin Chinese, stating that in “a wh-placeholder, the wh-word is typically preceded by the distal demonstrative marker *na* and the generic classifier *ge*” (p. 280). Apart from the fact that these complementary analyses demonstrate that the placeholder lies in the eyes of the beholder, they also indicate that it is probably the entire phrase that functions as a placeholder, rather than just one of its elements.

A similar issue concerns the Japanese adnominal demonstrative *ano*. This is categorized as a “demonstrative-derived placeholder” by Seraku (2022a: 120–121), yet in the examples it occurs together with semantically vague nouns, e.g. ‘thing’ or ‘place’. It is thus open to debate whether it is the demonstrative which carries the placeholder function or the vague noun, and it might be preferable to view the entire demonstrative-noun phrase as the placeholder, rather than just one of its elements.

Lastly, it should be noted that whereas it is relatively straightforward to identify a dedicated filler as a filler, it can be very difficult to distinguish between the source item and the filler in languages that still have both. We discuss this issue in more detail in §9.2.1.

4 A morphosyntactic typology of fillers

In this section, we outline our refined morphosyntactic typology of fillers, based on three criteria:

- the referentiality of the filler;

- the morphosyntactic integration of the filler;
- the possible part of speech of the syntactic projection.

The morphosyntactic integration and referentiality of the filler let us distinguish between placeholders and hesitatives, and whether or not the syntactic projection is restricted to particular parts of speech distinguishes between specific and general fillers. We therefore obtain a four-way distinction between SPECIFIC PLACEHOLDERS (§4.1), GENERAL PLACEHOLDERS (§4.2), SPECIFIC HESITATIVES⁶ (§4.4), and GENERAL HESITATIVES (§4.3). VERSATILE FILLERS (§4.5) are used for both placeholder and hesitative strategies; in their placeholder function they can be either general or specific.

It is important to note that many languages have several fillers, of either one or several types. In particular, Austronesian languages stand out typologically by their rich and varied repertoire of fillers. Thus, in Papuan Malay for example, apart from the placeholder *siapa* dedicated to substituting for personal names, the interrogatives *bagemana* ‘how’ and *apa* ‘what’ function as placeholders, and the demonstratives *ini* and *itu* function as placeholders for nouns and verbs as well as – albeit infrequently – as hesitatives (Kluge 2017: 288–297, 388–389).

4.1 Specific placeholders

Fillers that are always referential and morphosyntactically integrated are placeholders. Some placeholders are restricted to particular parts of speech or a particular subset of projected nouns or verbs and are hence SPECIFIC. Examples of placeholders that are restricted to particular parts of speech are the Georgian placeholder verb (Amiridze 2010) and the Komnzo placeholder *bäne/baf*, which is restricted to substituting for nouns and noun phrases (Döhler 2025 [this volume]), as exemplified in (3).

- (3) Komnzo (Döhler 2025 [this volume]: 290, ex.(13))
 Zöbthé zwa\wärez\é *bäne=me* (280ms) *kofä tot=me*.
 first 1SG>3SG.F:RPST:PFV/aim PH=INS fish spear=INS
 ‘First I aimed at it with the whatchamacallit... with the fish spear.’

Examples of placeholders that are restricted to a particular subset of projected nouns or verbs are the Jamsay placeholder *cě*: or *cì gé* ‘thing’ restricted to non-human nouns (Heath 2008: 475), and the Manambu placeholder verb

⁶Note that Dimock (2010: 136) already distinguishes between “syntactically specific delay fillers” and a “general interjection hesitator”.

məgi-, which is restricted to verbs of affect and process and cannot substitute for verbs of speech, emotion, or mental process, nor ditransitive verbs or stative verbs (Aikhenvald 2008: 575). Several languages have specific placeholders for names or terms for people (Vogel 2020). For example, in the Tungusic language Evenki there is a split between the placeholder *ajŋi/ajŋa*, which has a wide range of nominal and verbal targets, and the placeholder *uŋun*, which is restricted to replacing proper names, i.e. the names of humans, anthropomorphized animals, and places (Klyachko 2022: 214).

Sometimes a specific placeholder for verbs is derived from a specific placeholder for nouns. For example, the Algonquian language Maliseet-Passamaquoddy has a placeholder element *íy-* that is specific for nouns; from this, two specific verbal placeholders can be derived, namely *iy-i-* (animate intransitive, illustrated in (4)) and *iy-úw-* (animate transitive; LeSourd 2003). Similarly, in Dolgan the specific verbal placeholder *kimne:-* is morphologically derived from the specific nominal placeholder *kim*, itself identical to the interrogative pronoun ‘who’ (Däbritz 2018). In contrast, in Jamsay the verbal placeholder *cì ge kárná-* ‘do (a) thing’ is derived from the nominal placeholder with the help of a light verb construction (Heath 2008: 475).

- (4) Maliseet-Passamaquoddy (LeSourd 2003: 150)
Nékom 't-olömi=íy-i-n, 'kisahqé-wsa-n.
he 3-forward=PH-AI-SUB (3).uphill-walk-SUB
'He did something going forward, walked up the bank.'

4.2 General placeholders

Other placeholders, i.e. morphosyntactically integrated fillers, are GENERAL: they are polycategorial and can substitute for nouns, verbs, adjectives and even phrases. The Kolyma Yukaghir placeholder *xe-*, for example, can stand in for nouns, verbs, and demonstrative roots (Ventayol-Boada 2025 [this volume]), and the Indonesian proximal and distal demonstratives *ini* and *itu* (Wouk 2005) can replace nouns, verbs, and adjectives. Example (5) shows that the placeholder *itu* can have a verbal target. It then takes verbal morphology (here patient trigger and applicative).

- (5) Indonesian (Wouk 2005: 242)
Trus, ini-nya semua di-itu-in di-bilang-in.
then this-GEN all PT-PH-APPL PT-say-APPL
'Then everything (will) be that (will) be explained (to him).'

4.3 General hesitatives

We consider fillers that are never referential nor morphosyntactically integrated as hesitatives. These tend to be GENERAL, occurring in word searches of any kind or when the speaker is pondering how to continue their discourse. They can either be dedicated, like English *uh* and *uhm*, Nahavaq *a* (Dimock 2010), and Komnzo *ä* (Döhler 2025 [this volume]), or have a lexical source, like Korean *ku*, *ce* and *ceki*, which are also demonstratives in that language (Hayashi & Yoon 2006). Similarly, the Kalamang word *nain* ‘like’ functions as a general hesitative, as illustrated in example (6) (Visser 2025 [this volume]).

- (6) Kalamang (adapted from Visser 2025 [this volume]: 264, ex.(16))
Mu-nan nain opa nain neba-un me et kinkin=a saerak
 3PL-too HES ANA HES what-3POSS TOP canoe small=FOC NEG.EXIST
leng wa me.
 village PROX TOP
 ‘They too, like earlier, like whatsit, there are no small canoes in this
 village.’

4.4 Specific hesitatives

The definitions of placeholders and hesitatives found in the literature do not allow for hesitatives, i.e. fillers that are neither referential nor morphosyntactically integrated, that are SPECIFIC for certain types of delayed constituent. However, this type of filler is attested in Austronesian languages.

Nahavaq, a language spoken in Vanuatu, has numerous prefixal hesitatives that are “specific to the grammatical function of the word being retrieved” (Dimock 2010: 120) and that differ in form according to whether the delayed constituent is a personal name, a common noun, a location, or a verb (Dimock 2010: 121–127). For example, the hesitative *ni* is used when a speaker is trying to retrieve a common noun (7); this is formally related to the “nominal marker” prefixes *nV-* and *ni-* that occur with most nouns in the language and differs from the dedicated hesitative *a* or *ma* that occurs during a search for names or terms for people as well as from the hesitative *e* that indicates trouble with retrieving a location. The verbal hesitations are different yet again: they are identical to the prefixes that index subject and mood on verbs (8) and differ according to the person and mood of the elusive verb; they can even carry extra prefixes, such as the negative (9) or irrealis marker, which are found on the delayed constituent. They therefore convey information about the delayed constituent even while

the word search is going on, although they cannot be considered placeholders as such. They might at first glance resemble disfluent repetitions of the verbal prefix, but the filler always retains the vowel *e*, irrespective of the vowel of the prefix, while the vowel of the verbal subject-mood index assimilates to that of the base (10).

- (7) Nahavaq (adapted from Dimock 2010: 122)
En vales tuwan ko-log ke-vini ni na-lambut.
and time INDF 3SG.IRR-go 3SG.IRR-shoot HES NV-rat.
'And sometimes he would go and shoot a uhm rat.'
- (8) Nahavaq (adapted from Dimock 2010: 123)
... *gcen ndu ndu ndu-tig ha-haropw.*
because HES HES 1INCL.DU-roast DUP-quickly.over.flames
'...because we've uhm cooked it over flames.'
- (9) Nahavaq (adapted from Dimock 2010: 124)
I-noq re-vwer mi-s mi-s-makas veq, mi-koh tey.
3SG.R-like 3PL-say HES-NEG 1EXCL.PL-NEG-come.out NEG 1EXCL.PL-be FOC
'(You know) we didn't uhm come out, we just stayed.'
- (10) Nahavaq (adapted from Dimock 2010: 124)
Oveh, kinag ne no-log siley.
whoa 1SG HES 1SG.R-go far
'Whoa, I have uhm come from far away.'

The Austronesian language Patani spoken on Halmahera also has a set of subject-specific hesitatives. These have a status in between that of free subject pronouns and verbal subject prefixes and are used to gain time while a speaker searches for an elusive verb (Sjånes Rødvand 2024: 113–118). Additionally, in an article containing a preliminary analysis of hesitation phenomena in Western Subanon, a language of the Philippines, Blake (2020) describes three NP-specific hesitatives in addition to a general placeholder that can substitute for both nouns and verbs. The hesitatives are partially reduplicated forms of the syntactic pivot markers *og* 'pivot', *nog* 'relativizer' (11), and *sog* 'locative'.

- (11) Western Subanon (adapted from Blake 2020: 14)
Ongon dosop og laki; nogog pokpanow bu mama' nog lolingitan.
exist also PIV man HES walking and looks COMP angry
'There's also a man who-ah is walking and looks like (he's) angry.'

The syntactically specific hesitation markers in Nahavaq, Patani, and Western Subanon do not fit into the established typologies, which contrast morphosyntactically integrated placeholders that substitute for particular words with hesitatives that are not morphosyntactically integrated and function merely to stall for time. In contrast, these Austronesian forms carry grammatical meaning and are syntactically specific, but they do not hold the place for a particular word, they merely delay the completion of the utterance (cf. Blake 2020: 16). It can only be hoped that further research on more Austronesian languages will uncover more such cases, so that the extent of variation in these syntactically specific hesitatives can be uncovered.

4.5 Versatile fillers

In contrast to the fillers described above, some languages have a single VERSATILE FILLER, which is used both as a hesitative and a placeholder (either specific or general). This is found for example in Ilocano, a language spoken in the Philippines, where the “versatile empty root” *kua* is used as a hesitative, as in (12), as well as to substitute for nouns, verbs (13), or clauses (14) (Rubino 1996). A similar situation is found in the Tungusic language Negidal, where a single element functions as a general hesitative and a general placeholder (Pakendorf 2025 [this volume]).

- (12) Ilocano (adapted from Rubino 1996: 657)

<i>Ah, kua, bigla</i>	<i>nga n-ag-idda</i>	<i>ti</i>	<i>kawayan.</i>
HES HES	Suddenly LIG PST-INTR-lie:3SG.ABS ART		bamboo
'Ah, um, suddenly he lay down on the bamboo...'			
- (13) Ilocano (adapted from Rubino 1996: 659–660)

<i>N-ag-kua</i>	<i>didiey.</i>	<i>N-ag-waras-en</i>	<i>dayta nga</i>
PST-INTR-PH	DIST:DEM	PST-INTR-spread-COMPL:ASP	that LIG
<i>balita-n.</i>			
news-PART			
'That stuff <u>did</u> whatchamacallit, that news <u>spread</u> .'			
- (14) Ilocano (adapted from Rubino 1996: 658)

<i>Tatta n-ag-aramid-da</i>	<i>ngay ti</i>	<i>building nga kua-.. nga kasla</i>
now PST-INTR-made-3PL.ABS	PART ART	building LIG PH LIG like
<i>Shoemart</i>	<i>ti</i>	<i>itsora-na.</i>
Shoemart:mall	ART	appearance-3SG.ERG
'So then they made a building that is whatchamacallit.. that <u>looks like</u> Shoemart mall.'		

In other languages, such as the Nakh-Daghestanian languages Udi and Agul or Armenian, a single form can be used as a hesitative (15) as well as to substitute for nouns (16) and – in Armenian – even adjectives, while a verbal placeholder is derived from this form with a light verb (17) (Khurshudyan 2006: 81, Khurshudyan & Podlesskaya 2006, Ganenkov et al. 2010).

- (15) Armenian (adapted from Khurshudyan & Podlesskaya 2006: 7)

Na ban ibrev petkh a gar ēsor.

3SG HES as.if must AUX.PRS.3 come.PST.3 today

‘He, uhm, seems to have been supposed to come today.’

- (16) Armenian (adapted from Khurshudyan & Podlesskaya 2006: 9)

Hiš-um es ban-ə ber-el ēir aparat-ə...

remember-IPFV AUX.2SG PH-DEF bring-PFV AUX.PST.2SG camera-DEF

‘Do you remember, you brought the whatchamacallit, the camera, ...’

- (17) Armenian (adapted from Khurshudyan & Podlesskaya 2006: 13)

Yes ēl mi angam ban ar-echi gn-achi ēd sayth-ə

1SG also one time PH do-AOR.1SG go-AOR.1SG this site-DEF

spyware-i...

spyware-GEN

‘I also once **did** whatchamacallit, went onto this site “spyware”...’

Surprisingly, such versatile fillers are not uncommon in the world’s languages: apart from the languages mentioned above, they are found in Even and Evenki (Matić 2008, Klyachko 2022), Mandarin Chinese (Hayashi & Yoon 2006), Tagalog (Nagaya 2022), Nasal (Billings & McDonnell 2025 [this volume]), Besemah (McDonnell & Billings 2025 [this volume]), Papuan Malay (Kluge 2017: 388–389), Manambu (Aikhenvald 2008: 573–574), Amazonian Spanish (Vallejos Yopán 2023), Northern Pastaza Kichwa (Rice 2025 [this volume]), Sliammon Salish (Watanabe 2010) and French (Hennecke & Mihatsch 2022), and further languages listed in Seraku (2025). It is possible that versatile fillers have been overlooked by researchers biased by the terminological dichotomy of placeholders vs. hesitatives, and that close examination of naturally occurring fillers will reveal even more examples.

5 Extended functions of fillers

Fillers are mainly discussed in the light of disfluencies, e.g. in Hayashi & Yoon (2006), who focus on ‘demonstratives as ‘filler words’ in contexts where speakers

encounter trouble recalling a word or selecting the best word to use to designate some entity during the course of producing an utterance" (485, our highlighting). Similarly, the paper by Podlesskaya (2010) "focuses on a type of discourse marker that signals production difficulties in spontaneous spoken discourse" (11, our highlighting). However, it was pointed out early on (e.g. Enfield 2003) that place-holders can be used intentionally by speakers, with a variety of functions, namely socially motivated avoidance of a particular term (§5.1), to manage interactions (§5.2), and to refer to arbitrary referents (§5.3). However, it should be noted that in studies based on oral corpora it is not always clear what motivates particular instances of filler use.

5.1 Socially motivated functions

Socially motivated intentional uses have been given different labels in the literature: "therapeutic" vs. "diplomatic" (Tárynyiková 2019), "communicative" vs. "social" (Seraku 2020), and " PH_A " for placeholders motivated by the speakers' ability vs. " PH_P " for uses motivated by the speakers' preference (Seraku 2024). Among the socially motivated intentional uses, two major sub-functions can be distinguished, namely the "avoidance" of socially sensitive terms (aka taboo) and the "conspiratorial" function, used "to prevent potentially overhearing third parties from understanding, and/or to create a collusive air between interlocutors" (Enfield 2003: 106, see also Hayashi & Yoon 2006: 501–507, Keevallik 2010, Cheung 2015). Example (18) shows the use of a placeholder to avoid embarrassment due to the explicitness of 'die'. Example (19) shows the use of a placeholder to prevent dinner guests from understanding the hosts' plan for a surprise.

- (18) Mandarin Chinese (adapted from Cheung 2015: 276)

Na ge lao taitai yijing na-ge shenme-le.

DEM CLF old lady already DEM-CLF PH-PFV

'The old lady has already ... you-know-what-ed (= died).'

- (19) English (Enfield 2003: 106)

[After dinner, the host says to his wife] I think it's time to serve the you-know-whats (=peaches).

Palacios Martínez & Núñez Pertejo (2015) list several pragmatic meanings of English placeholders, such as when they are used derogatorily, as insults, with the goal of not sounding pretentious, as euphemisms, and informally to build in-group identity. A further intentional strategy of placeholders are "rhetorical" uses, found for instance with Japanese demonstrative-derived placeholders used

in internet articles and blog posts: these are written not in the *hiragana* script standardly used for demonstratives, but in the distinct *katakana* script, thus visibly signalling their special function to the reader (Seraku 2022a). Furthermore, it is possible that in the Tupí-Guaraní language Teko a particular construction involving the filler has a euphemistic or suspense-creating function (Rose 2025 [this volume]); however, this cannot be asserted with certainty due to the small number of examples.

Placeholders can be functionally specific, with forms used only in contexts of disfluency, such as *urjun* in Negidal (Pakendorf 2025 [this volume]), or forms used only in socially motivated contexts, where the target is mostly omitted, such as Nahavaq *na-lan* (Dimock 2010: 130–131). In several languages, both types of functionally specific placeholders co-exist. For example, in Kuuk Thaayorre, a Pama-Nyungan language of northern Australia, the distal demonstrative *yuunhul* serves as a filler in situations where the speaker wants to “signal that the inaccessibility of the target lexeme [...] is disruptive to the flow of speech, and must be repaired before proceeding” (Gaby 2017: 253), whereas the proximal demonstrative *inhul* is used when the speaker feels that it is unnecessary or even undesirable to fill in the target, such as in taboo situations (Gaby 2017: 254). English makes a similar distinction between *whatchamacallit*, used when a speaker cannot retrieve or doesn’t know the target word, and *you-know-what*, used when the speaker does not want to utter the target for particular social reasons (Enfield 2003: 105–107). Similarly, Kalamang has distinct placeholders used for word retrieval and taboo substitution (Visser 2025 [this volume]). Thus, in these languages the distinction between “ability” and “preference” (Seraku 2024) is formally marked. Such functional specificity is far from being generalized, however: speakers of the Papuan language Komnzo and the Austronesian language Nasal use the same filler in taboo situations as in situations of word search (Döhler 2025 [this volume], Billings & McDonnell 2025 [this volume]).

5.2 Interaction management

Several studies have demonstrated that fillers (both hesitations and placeholders) are used not only in contexts of word search or trouble with speech planning, but also to manage interactions in discourse (e.g. Huang & Tanangkingsing 2005, Keevallik 2010: 162–167, Watanabe 2010: 178, Jehoul et al. 2016, Kosmala & Crible 2022). Clark & Fox Tree (2002: 90) list many interpretations of the English fillers that are related to interaction management, like “speakers want to keep the floor”, “speakers want to cede the floor”, “speakers want the next turn”, among others.

Allwood et al. (2005) therefore prefer to use the term “Own Communication Management” to cover phenomena otherwise referred to as “hesitation”, “disfluency”, or “self-repair”, and Kosmala & Crible (2022: 220) introduce the term “fluenceme” for elements like fillers, in order to underline their “potential to serve both fluent and disfluent functions”. Example (20) shows how a filler (the Estonian pronominal demonstrative *see*) is used turn-initially, here specifically to initiate a “reason-for-the-call turn”.

- (20) Estonian (adapted from Keevallik 2010: 164)⁷
 =*see*, .*h* *ma tahsin seda küsida et*
 FIL I want:PST:1SG this:PRT ask:INF that
 ‘Uhm, I wanted to ask you’

5.3 Arbitrary reference

Even though it has been posited that the prototypical use of a placeholder is to signal to the hearer that a specific referent should be looked for in the context (e.g. Hennecke & Mihatsch 2022: 300), Seraku (2022b) identifies arbitrary reference as a function of Japanese and Korean interrogative-derived and Romanian and Bulgarian demonstrative-derived placeholders (21). In such uses, the placeholder is not used to substitute for a specific referent, but precisely to fill in for some arbitrary entity.

- (21) Japanese (Seraku 2022b: 436)
Boku jitsuwa mainichi nikki kai-teru-ndesu. Kyoo-wa nani-ga
 1SG in.fact every.day diary write-IPFV-MM.HON today-TOP PH-NOM
at-ta-toka nanishi-ta-toka nani tabe-ta-toka.
 happen-PST-etc. PH-PST-etc. PH eat-PST-etc.
 ‘[In this post, the writer conveys that writing up a diary every day reduces his stress.] In fact, I write in my diary every day, like “Such-and-such happened today,” “I did such-and-such,” “I ate such-and-such,” and so on.’

5.4 Issues with extended functions

This section has presented several functions of fillers beyond their use in disfluency. It should be noted, however, that the extended functions of placeholders

⁷The equal sign in the text line indicates “latching of turns or words”, and .*h* indicates in-drawn breath (Keevallik 2010: 171).

(avoidance and conspiratorial use, interaction management or arbitrary reference) have not been identified for dedicated placeholders with opaque etymology, but only for placeholders derived from demonstratives, interrogative pronouns, or general/semantically empty nouns. This can partly be explained by the fact that certain functions have only recently been identified, such as the rhetorical or the arbitrary reference function (Seraku 2022a, 2022b). It is thus possible that some dedicated fillers might be found to have these functions if more attention is directed to finding them. Nevertheless, the paucity of dedicated filler items with extended functions raises the question whether these extensions are actually functions of the filler, or whether they aren't rather functions of the base form, i.e. the demonstrative, interrogative, or general noun, which developed into a filler on the one hand and into a marker of arbitrary reference or rhetorical flag on the other.

6 Fillers and general extenders

Several descriptions of fillers include general extenders in the scope of their discussion (e.g. Aikhenvald 2008: 573–575, Ganenkov et al. 2010, Maisak 2023, Klyachko 2025 [this volume], Rose 2025 [this volume]). Such general extender expressions add an item with non-specific reference to an existing (even minimal) list, thus creating an “ad hoc category” (Mauri & Sansò 2018). They are optional structures which typically consist of a conjunction and a noun phrase, occur in phrase- or clause-final position (Overstreet & Yule 2021: 1), and can be translated as ‘and so on’ or ‘et cetera’. What fillers, and notably placeholders, and general extenders have in common is that they both belong to the category of vague language. In addition, both fillers and general extenders frequently develop out of general nouns or interrogatives (Hayashi & Yoon 2006; Mauri & Sansò 2018: 28), and are therefore formally related. For instance, the *baʔe* root in Teko is used as a noun for non-humans (22), as a hesitative (23), and as a general extender (24).

- (22) Teko (Rose 2025 [this volume]: 449, ex.(4))
O-ho-pa baʔe-kom-a=nam o-apig=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]
- (23) Teko (Rose 2025 [this volume]: 458, ex.(30))
Kob (0.3) pitaj-am (2.3) baʔe (1.2) kito-r=ehe e-iba.
EXIST child-TRANSF HES frog-RELN=with 3-pet
'There is a child with his um... pet frog.' [13.001]

- (24) Teko (Rose 2025 [this volume]: 461, ex.(36))
Dati arakapusa (0.8), dati fort, t-iru, ba?e-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 diachronic link between fillers and general extenders remains to be investigated; three scenarios are found in the literature. Firstly, the general extender could have developed out of the filler, as argued for Udi and Agul by Ganenkov et al. (2010: 111–114). Secondly, the general extender may have been the base from which the filler developed. For example, Klyachko (2025 [this volume]) suggests that in some Tungusic languages the interrogative-based placeholders developed under the influence of the use of the interrogative stem as a general extender found in the entire family. Thirdly, it is possible that the source element developed separately into a filler and into a general extender, as with the extended functions of fillers discussed in §5.4; this is the analysis provided by Rose (2025 [this volume]) for Teko.

7 Fillers in language contact

Very little is known about the impact of language contact on the use of fillers. Data from Spanish-English bilinguals in Boston show that the age of arrival in the US and the amount of Spanish-exclusive communication have an impact on the phonetic shape of the hesitative *uh*: individuals who arrived before late adolescence (~15 years of age) or who were born in the US and who have relatively few interlocutors with whom they use exclusively Spanish are far more likely to use a central vowel ([a] or [ə]) as their hesitative when speaking Spanish than the form commonly found in Spanish, namely [e] (Erker & Vidal-Covas 2022). This difference is likely to be due to the influence of English and demonstrates that language contact can have an impact on the form of the hesitative. Spanish borrowed a placeholder with arbitrary reference, *fulano/fulana*, from Arabic, with earliest attestations from the Middle Ages (Gerhalter & Salaaoui 2020). In both languages the placeholder was predominantly used in juridical and ritual texts where it replaced proper nouns, such as in prescriptions as to what the groom and bride were to say during a marriage ceremony.

Borrowed fillers are further mentioned for Indonesian (placeholder *anu*, from Javanese; Williams 2009: 9), Kalamang (filler *apa* borrowed from Indonesian; Visser 2025 [this volume]), Kolyma Yukaghirs and Tungusic languages (demonstrative-derived filler *èto* borrowed from Russian; Ventayol-Boada 2025

[this volume], and Klyachko 2025 [this volume], respectively), and Northern Pastaza Kichwa (demonstrative-derived hesitative *este* from Spanish; Rice 2025 [this volume]). In Besemah, contact influence from Jakarta Indonesian might have increased the frequency of use of the demonstrative-derived filler *ini* over the dedicated filler *anu* (McDonnell & Billings 2025 [this volume]). However, while these brief descriptions mention the presence of these elements, which were mostly borrowed from the sociopolitically dominant language, they do not provide information on the relative frequency of the borrowed vs. the indigenous forms,⁸ and they also leave open the question to what extent these fillers of foreign origin are really integrated into the language as opposed to their occurring as nonce borrowings or in code-switches. It is thus clear that there is still a lot of scope for research in the domain of filler forms and functions in language contact situations.

8 Fillers and the nature of language

Fillers are not only important from a descriptive point of view, but also for their potential to enlighten us on how speech is produced and how it functions in interaction. The fact that placeholders mirror the morphology of their targets, such as both case and evidential markers in Northern Pastaza Kichwa (Rice 2025 [this volume]), aspect and mood plus subject marking in Evenki, as illustrated in (25), or case and possessive marking in Negidal (Pakendorf 2025 [this volume]), provides evidence for the fact that the syntactic structure of an utterance is in place before the target lexeme is accessed.

- (25) Evenki (adapted from Klyachko 2022: 209)

<i>D'əm-mu:-l-mi</i>	<i>ayi-ŋna-kal</i>	<i>gu:-sə:</i>
eat-DES-INCH-CVCOND	PH-HAB-IMP.2SG	say-PANT
<i>əri-ŋ-mə-w</i>	<i>tuge: s'iwu-ŋna-kal.</i>	
this-INDR.POSS-ACC-1SG.POSS	so	lick-HAB-IMP.2SG

'If you get hungry, he said, do that thing, lick this one (paw) of mine.'

Mismatches between the placeholder and the target also demonstrate that the syntactic specifications of an utterance are put in place early on. Thus, in Negidal mismatches between case morphemes on the placeholder and the target tend to

⁸An exception is the presentation by Egorova et al. (2021) on borrowed fillers in Evenki, who mention that in oral recordings native fillers are preferred over borrowed ones, even by semi-speakers (with indigenous fillers occurring more than 4.5 times as often as borrowed ones).

be semantically congruent, involving cases that can both be used to mark goals or direct objects, for example (Pakendorf 2025 [this volume]). Similarly, the Russian complex placeholder *ètot... kak ego* (literally ‘this one how is it (called)’, or ‘what is its (name)’) carries the same gender and number as the target.⁹ However, in some cases there is a mismatch in gender between the placeholder and the target, such as in (26), where the placeholder carries feminine gender (also found on the modifier *gosudarstvennuju* ‘State’), but the target is neutral gender. Here, it is probable that the initial target of the placeholder was the feminine-gender word *premija* ‘bonus’, which was replaced with the final target *voznagraždenie* ‘reward’. This shows that in speech production the grammatical specifications of an item, such as grammatical gender, are accessed independently and in advance of the actual lexeme (Podlesskaya & Korotaev 2022: 68–69).

- (26) Russian (adapted from Podlesskaya & Korotaev 2022: 68)
- | | |
|------------------------------------|---|
| <i>Nu ladno poluciš'</i> | <i>gosudarstvennuju... èto... kak eë...</i> |
| well OK | you.will.get State[ADJ.F] |
| <u><i>voznagraždenie</i></u> [...] | PH[F] |
| reward[N] | |
- ‘Well, OK, you will receive a State whatchamacallit, reward...’

9 Scope and contribution of the volume

This section first presents the scope of the volume in terms of the languages discussed (9.1), then discusses major issues (9.2) and emerging topics (9.3). In our discussion, we restrict ourselves to the languages of the volume.

9.1 Languages discussed in the volume

As summarized in Table 2, fillers in 16 languages are discussed in this volume, with 10 chapters being devoted to individual languages and the chapter by Klyachko providing a comparison of fillers and general extenders in the Tungusic family as a whole. The table lists the languages in geographical order, from west to east and north to south. This is also the order of inclusion in the volume.

The languages represented in this volume are found in different macro-areas (Figure 1). Paponesia is the best represented, with descriptions of fillers in both Austronesian and Papuan languages, while African languages are unfortunately

⁹The first element, the proximal demonstrative *ètot~èta~èto* also takes the case of the target; it can, however, also stand in the invariant neuter gender form *èto*.

Table 2: Languages discussed in the present volume

Language	Glottocode	Macro-area	Family	Author
Evenki	even1259	Eurasia	Tungusic	Klyachko
Even	even1260	Eurasia	Tungusic	Klyachko
Udihe	udih1248	Eurasia	Tungusic	Klyachko
Oroch	oroc1248	Eurasia	Tungusic	Klyachko
Nanai	nana1257	Eurasia	Tungusic	Klyachko
Kur-Urmi Nanai	kuro1242	Eurasia	Tungusic	Klyachko
Negidal	negi1245	Eurasia	Tungusic	Pakendorf, Klyachko
Kolyma Yukaghir	sout2750	Eurasia	Yukaghir	Ventayol- Boada
Besemah	cent2053	Papunesia	Austronesian	McDonnell & Billings
Nasal	nasa1239	Papunesia	Austronesian	Billings & McDonnell
Kalamang	kara1499	Papunesia	West Bomberai	Visser
Komnzo	komn1238	Papunesia	Yam	Döhler
Dalabon	ngal1292	Australia	Gunwinyguan	Ponsonnet
Mohawk	moha1259	N. America	Iroquoian	Mithun
Northern Pastaza Kichwa	nort2973	S. America	Quechuan	Rice
Teko	emer1243	S. America	Tupian	Rose

not included at all. Thus the volume unintentionally mirrors the general state of cross-linguistic descriptions (§2).

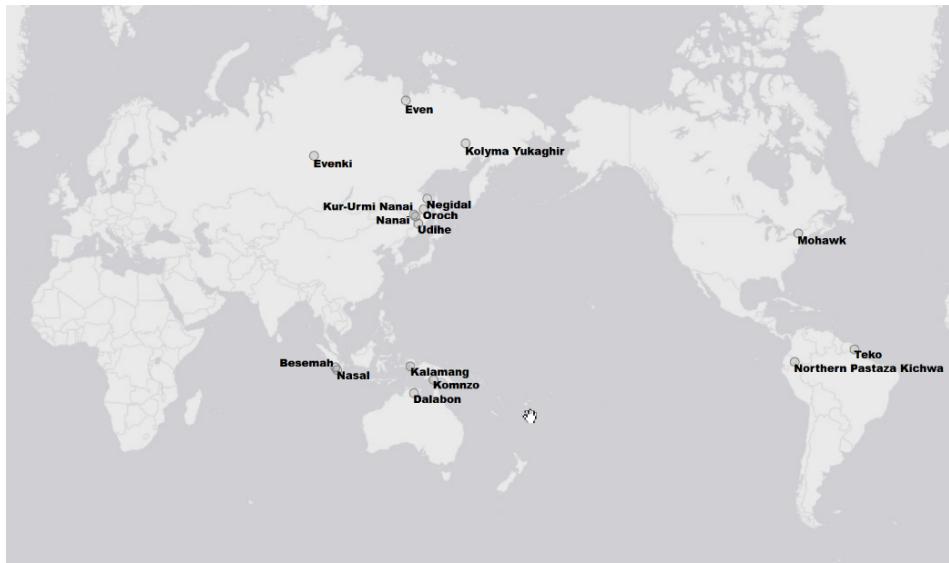


Figure 1: Map of the languages studied in the volume

All studies are based on (mostly first-hand) corpora of oral recordings, sometimes supplemented with video recordings which allowed the authors to include observations on the gestures that accompany particular uses of fillers (see §9.3.3). While most chapters provide synchronic descriptions of fillers, three chapters deal with diachronic aspects to various extents: Ventayol-Boada suggests that the Kolyma Yukaghir placeholder may have developed out of the copula *λe-*, and Rice proposes that the Northern Pastaza Kichwa filler *mashti* developed out of the phrase “what name”, with the placeholder uses developing first via phonetic erosion and semantic bleaching. The placeholder later evolved into both a hesitative (via cooptation) and a pro-verb (a case of (re)lexicalization). Rose discusses the development of discourse functions of the general noun with non-human reference *baʔe* in Teko: the general noun probably developed into a hesitative via placeholder uses, in a process that involved semantic bleaching, reanalysis, extension, and prosodic changes. The hesitative may have been the source for the general verb ‘do’, though that remains somewhat speculative due to the lack of bridging contexts in the corpus. In contrast, the general noun independently developed interrogative, general extender, rhetoric and nominalization functions.

9.2 Major issues

This section highlights four major issues identified on the basis of the chapters within this volume: 1) the distinction between fillers and related elements; 2) the distinction between the hesitative and the placeholder uses of a given filler; 3) the rather low frequency of overt targets after placeholders in several languages; and 4) individual speaker variation.

9.2.1 Methodological issues with identifying fillers

Dedicated fillers can straightforwardly be identified as such; however, disentangling filler uses from uses of the source item (i.e. demonstrative, interrogative, or general noun) or from future developments¹⁰ can be problematic in languages where fillers co-exist with their sources or their extensions. Criteria used by the authors of the chapters in the present volume to distinguish fillers from their sources are the following:

- form
- prosody
- position within utterances
- function in the discourse structure
- semantics
- morphological combinability
- syntactic distribution

The chapter by Rose, for example, shows how the syntactic, morphological, semantic and prosodic characteristics of *baʔe* in Teko are sufficient to distinguish three lexical units and identify particular occurrences as either nouns, verbs or hesitatives. As a noun for non-human referents, *baʔe* is found in the syntactic position of nominals, takes nominal morphology, refers to non-human entities, and is fully integrated in the phrase in which it occurs. As a verb, *baʔe* is found in the syntactic position of verbs and takes verbal morphology. As a hesitative, *baʔe* does not show a restricted syntactic distribution, it never takes any morphology,

¹⁰In this volume, fillers are described as developing into predicates (verb ‘make’ in Teko and pro-verb in Mashti), and connectors (in Komnzo) or markers of dependency between clauses (in Mohawk).

can be found before delayed constituents referring to human entities, and is often lengthened and surrounded by pauses.

However, there are limits to the use of formal criteria to distinguish fillers from their sources or their future developments. For example, in Teko, the identification of placeholder uses of the same form *ba?e*, distinct from the basic general noun, is not straightforward. The reverse might also be true: Döhler explains how in a first analysis of Komnzo *bäne* he considered it to be both a demonstrative and a placeholder. However, a careful examination of all text examples leads him to the conclusion that all occurrences are actually linked to disfluency, and he does not consider *bäne* as a demonstrative any more. In Mohawk, the proximal and distal demonstratives are used both in contexts of disfluency as hesitations and especially as placeholders, but they are also found in contexts without disfluency: first, in a discourse structure where they basically allow the speakers to hold the floor, and second in a complex syntactic structure, marking the fact that a complement or relative clause follows. However, the different uses cannot be distinguished by their prosody, their morphology (they do not take any), or their position. In the lack of formal evidence for the speaker's communicative intention, the analysis is based solely on the investigator's interpretation of the discourse function.

9.2.2 Indeterminacy of placeholder and hesitant use in versatile fillers

Hesitations and placeholders are often portrayed as being separate elements that are clearly distinguishable (most notably in Hayashi & Yoon 2006). This is also found by Visser for Kalamang, where hesitations and placeholders are distinct lexical items with no functional overlap. In other languages the filler can straightforwardly be analysed as being mostly a hesitative, as in Teko (Rose), or a placeholder, as in Kolyma Yukaghir (Ventayol-Boada), Dalabon (Ponsonnet), and Mohawk (Mithun). However, there are also languages with versatile fillers which are used both as hesitations and placeholders (§4.5), namely Negidal (Pakendorf), Besemah (McDonnell & Billings), Nasal (Billings & McDonnell), and Northern Pastaza Kichwa (Rice). The relative frequency of one or the other strategy differs across the languages: while in Nasal the demonstrative-derived fillers predominantly function as placeholders and the interrogative mainly functions as a hesitative, in Besemah placeholder uses dominate for all filler forms. Similarly, in Negidal placeholder uses are about twice as frequent as hesitative uses; in contrast, in Northern Pastaza Kichwa the hesitative use is by far the most frequent in the corpus, even though it can be assumed to have developed out of the placeholder use.

What is striking, however, is that in languages with versatile fillers it is at times hard or even impossible to determine what function a particular token of the filler might have, in spite of analyses of prosody and intonation. These studies thus demonstrate that in some languages the filler is indeed a single polyfunctional item with a continuum of uses from hesitant to placeholder. Whether this indeterminacy also holds for the other languages discussed in §4.5 would need to be verified with prosodic analyses.

9.2.3 Low frequency of overt targets

A third striking result is that targets are regularly not present after a placeholder in several languages discussed in the volume, namely in Negidal (Pakendorf), Kolyma Yukaghir (Ventayol-Boada), Besemah (McDonnell & Billings), Nasal (Billings & McDonnell), Kalamang (Visser), and Dalabon (Ponsonnet). For instance, the Negidal placeholder is followed by an overt target in 70% of its occurrences as a nominal placeholder, but in only 50% of its occurrences as a verbal placeholder. Even more surprising, in Besemah and Kalamang targets are actually more frequently absent than present after a placeholder. McDonnell & Billings indicate that the Besemah placeholders are followed by a target (a repair, in their terminology), in only one third of all occurrences.

This is an important finding, since there is a bias in the literature towards examples with overt targets for obvious expository reasons (openly acknowledged by Dimock 2010). This bias generates an implicit expectation that a placeholder will of necessity be followed by its target – a situation that is explicit in the terminological choice made by Khurshudyan & Podlesskaya (2006), who distinguish between placeholders (*preparativnaja podstanovka*, literally “preparatory replacement”), which are followed by their target, and “approximate nominalizations” (*priblizitel'naja nominalizacija*), which lack an overt target. The corpus-based studies included in the volume, with meticulous coding of each occurrence of the placeholders in a given language, thus provide important insights into the actual nature of placeholders. For instance, it is possible that placeholders do not always function as substitutes of specific targets that elude the speaker (and hence do not necessarily have the function to “signal that the speaker is not able or willing to provide a more specific target expression” (Hennecke & Mihatsch 2022: 300)), but that these semantically vague expressions maintain or develop some referential uses.

9.2.4 Variation between speakers

Finally, several studies in the volume (Pakendorf, Ventayol-Boada, McDonnell & Billings, Visser, and Ponsonnet) find notable differences in the frequency of use of fillers between individual speakers of a language, as also found for hesitatives among German women (Braun et al. 2023) and placeholders among English speakers of various ages (Palacios Martínez & Núñez Pertejo 2015). Such differences among speakers can partly be explained by their sociolinguistic profile: McDonnell & Billings show that younger speakers of Besemah, and especially those who have spent time outside of the region, strongly prefer the demonstrative pronoun *ini* over the dedicated filler *anu*. The authors explain this preference by contact with Jakarta Indonesian, which employs demonstrative pronouns as fillers. In contrast, Ponsonnet finds that it is individual speakers' preferences, which she calls 'styles', that determine placeholder use in Dalabon, with one speaker favouring lexical accuracy and hence using the placeholder to substitute for a (mostly nominal) target that is subsequently supplied, while another speaker favours fluidity of speech and therefore does not necessarily supply the target, with the placeholder standing in with nearly equal frequency for verbs and nouns. These styles are not strictly correlated with the speakers' proficiency. Pakendorf also suggests that the differences in the frequency of use of the Negidal filler *uŋjun*, as well as the differences in choice between hesitatives and placeholders, can only be partly explained by the proficiency of the speakers: while a semi-speaker of Negidal indeed uses it most frequently, a very proficient speaker also uses fillers very often (see §9.3.4 for other explanations).

9.3 Emerging topics

This section broaches several topics that have to date been rarely discussed in the literature, but which are touched upon in the studies included in the volume, and which we consider worth exploring in greater detail and in diverse languages in the future: the interaction of fillers 1) with prosody, 2) with other markers of disfluency or other fillers, 3) with gestures, and 4) the frequency of fillers in discourse as well as 5) the impact of language contact on fillers.

9.3.1 Fillers and prosody

Most chapters in the volume include some discussion of prosody, a topic with little prior coverage in the cross-linguistic literature on fillers (but see Dimock 2010, Podlesskaya & Korotaev 2022, Hennecke & Mihatsch 2022, and Vallejos Yopán 2023). In the chapters of the volume, the discussion mostly focuses on

duration (i.e. lengthening of fillers) and pauses, but sometimes also deals with prosodic contours. The prosodic studies may have two different goals. Firstly, some authors use prosody to distinguish fillers from homonymous forms. Rose uses prosodic features to investigate the distinction between the Teko hesitative and its source noun *ba?e*. The hesitative is prosodically more salient via final lengthening, a higher frequency of pauses preceding and following it, and longer duration of following pauses. In contrast, Ventayol-Boada does not find pauses to be relevant for distinguishing the filler and copula uses of *λe-* in Kolyma Yuktghir; rather, it is the position of the filler and the number of words in the intonation unit as well as the intonation contour that discriminate the two uses. In Komnzo, the close examination of the prosody of all occurrences of *bäne* leads Döhler to the conclusion that it is not a proper demonstrative in synchrony anymore, but always a placeholder. Secondly, some authors investigate how prosody correlates with different uses of the same filler. Rice compares the features of prosodic lengthening and pauses in relation to the different uses of the Northern Pastaza Kichwa *mashti* element, namely as a hesitative, a placeholder, and a pro-verb. In general, the hesitative use attracts more pauses and prosodic lengthening. McDonnell & Billings also show that hesitative uses of Besemah fillers are more frequently associated with other disfluency cues than their placeholder uses. In contrast, Pakendorf and Mithun do not find any correlation between prosodic patterns and uses of fillers as hesitations or placeholders in Negidal and Mohawk, respectively, mirroring the lack of distinction between the hesitative and placeholder strategies of fillers in French *truc* and *machin* (Hennecke & Mihatsh 2022). The lack of a cross-linguistically clear pattern concerning the role of prosody with respect to fillers motivates our call for further research on that question.

9.3.2 Fillers and other markers of disfluency

Several authors (McDonnell & Billings, Döhler, Rice) also touch upon other markers that accompany hesitations, essentially non-lexical ones like false starts, glottalization, and repetitions, in addition to pauses and lengthening. Furthermore, some authors mention the existence of other fillers in the language they study (Visser, Döhler, Rice). How the different lexical fillers and other types of filled pauses combine or complement each other is worth being investigated in more detail in more languages.

9.3.3 Fillers and co-gestures

Three chapters (Pakendorf, Döhler, Rice) include some discussion of gestures associated with fillers, a topic rarely approached for languages other than English (but see Hayashi 2003 on Japanese, Navarretta 2016 on Danish, and Graziano & Gullberg 2018 and Kosmala 2024 for comparative studies). Rice offers a systematic study of three types of gestures accompanying the Northern Pastaza Kichwa filler *mashti*: gaze aversion, excessive blinking and manual gestures. Manual gestures can themselves undergo disfluency, such as being “on hold”, repeated, or corrected. In general, it seems that gaze aversion, manual gestures, and gestural disfluency are often associated with the filler. Just like prosodic cues, excessive blinking and gestural disfluency occur more with the hesitative than with the placeholder use of *mashti*. Döhler finds that about two thirds of the Komnzo placeholders are accompanied by a gesture, either hand gestures, or lip- or head-pointing gestures, which thus provide a “parallel support channel”, helping the speaker find and the hearer identify the target. Hand gestures are the most common, often pointing to the referent (in line with the demonstrative origin of the placeholder). Other gestures identify a visible target or re-enact it. In Negidal, Pakendorf finds the same type of gestures accompanying the versatile filler *ujun* as those found in Northern Pastaza Kichwa and in Komnzo: gaze aversion, pointing gestures, and re-enacting gestures. She calls for a systematic study of gestures in future investigations of fillers in the languages of the world, with a close look at the timing of the gesture and the utterance of the filler, and the different possible functions of the filler.

9.3.4 Frequency of fillers

As much as possible, the authors of the chapters included frequency counts of the fillers under study in their corpus (remember that all studies in the volume are corpus-based). The frequency of fillers can vary depending on various parameters, related to the speaker (§9.2.4), content and context of the utterance, as well as the speech situation (Corley & Stewart 2008). For example, differences in the frequency of occurrences of fillers can partly be explained by genre and recording situation. Thus, Visser finds that placeholders are more common in conversations than in narratives, probably because the former are less planned. Also, Visser interprets differences in the individual speakers' uses of fillers in her Kalamang corpus as being mainly related to the topic of the conversation: one speaker used very many fillers in a recording in which he was urged to talk about herbal medicine and where he had to search his memory for names

of plants. Similarly, Pakendorf finds that an excellent Negidal speaker uses the filler very frequently while telling traditional fairy tales; here, it might be the cultural pressure to tell such tales in a particular manner that led him to hesitate frequently.

A more general question is whether there are cultural differences in the overall frequency of fillers. It is striking that in all the chapters of the volume, the frequency of fillers in the various languages is generally higher than that found in preceding studies (Zhao & Jurafsky 2005, Podlesskaya & Kibrik 2009). This high frequency of use of fillers found in typologically and geographically disparate languages indicates that they are far from being the marginal phenomenon they have been considered to date, leading us to call for more dedicated studies on fillers in the languages of the world.

9.3.5 Borrowed fillers

Finally, as briefly discussed in §7, there is sporadic mention of borrowed fillers in some chapters included in the volume (Klyachko, Ventayol-Boada, Visser, Rice). However, detailed studies of the effect of language contact on the form, frequency, and usage patterns of fillers are still missing. Given that discourse markers such as “fillers, tags, interjections, and hesitation markers” are among the most contact-sensitive elements of language (Matras 2009: 193), this would constitute a fruitful field for further research.

Acknowledgements

Some of the ideas developed here were presented at the workshop on “Placeholders in East and West” at the University of Tübingen in October 2024; we thank the audience for their questions. We also thank Martin Haspelmath, Eline Visser and Albert Ventayol-Boada for reading and commenting on the paper, and most especially Wiltrud Mihatsch for her expert feedback. Needless to say, any errors are entirely our responsibility.

We are furthermore extremely grateful to Rémi Anselme for his assistance with eliminating the numerous formatting and cross-referencing errors that arose as a result of conversion of the initial Word file to L^AT_EX, and thank Sébastien Flavier for creating the map in Figure 1.

Abbreviations

This list includes only the glosses which are not found in the Leipzig Glossing Rules.

AI	animate intransitive	NSP	non-specific possessor
ANA	anaphoric demonstrative	NV	nV- nominal prefix
AOR	aorist	PANT	anterior participle
ASP	aspect	PART	particle
COMPL	completive	PART.O	partitive object
CONT	continuative	PF	particle-final
CVCOND	conditional converb	PH	placeholder
DES	desiderative	PIMP	past imperfective
DUP	reduplication	PIV	pivot
EXIST	existential	PRT	partitive
FIL	filler	PT	patient trigger
HAB	habitual	R	realis
HES	hesitative	REF	referential
HON	honorific	RELN	relational
INCH	inchoative	RPST	recent past
INDR.POSS	indirect possession	SEQ	sequential
LIG	ligature	SUB	subordinative
MM	modal marker	TRANSF	transfer

References

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Allwood, Jens, Elisabeth Ahlsén, Johan Lund & Johanna Sundqvist. 2005. Multi-modality in own communication management. In *Proceedings from the second Nordic conference on multimodal communication*, 1–20. Göteborg: Göteborg University.
- Amiridze, Nino. 2010. Placeholder verbs in Modern Georgian. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 67–94. Amsterdam: John Benjamins.
- Amiridze, Nino, Boyd H. Davis & Margaret Maclagan (eds.). 2010. *Fillers, pauses and placeholders* (Typological Studies in Language 93). Amsterdam: John Benjamins.

- Bajo Pérez, Elena. 2019. El antropónimo sustitutivo. Ejemplificar, reconstruir, recrear, evocar: *Fulano,-a* y sus variantes y equivalentes [The substitute anthroponym. Exemplify, reconstruct, recreate, evoke: *Fulano,-a* and its variants and equivalents.] *Moenia: Revista license de lingüística & literatura* 25. 101–147.
- Billings, Blaine & Bradley McDonnell. 2025. Fillers in the gray: Indeterminacy in their use in everyday conversation. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 211–250. Berlin: Language Science Press. DOI: ??.
- Blake, A. L. 2020. Delays without fillers: The example of Western Subanon. *University of Hawai'i at Mānoa Working Papers in Linguistics* 51(1). 1–25. <http://hdl.handle.net/10125/73275>.
- Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt & Company.
- Blust, Robert. 2013. *The Austronesian languages*. Canberra: Asia-Pacific Linguistics.
- Blust, Robert & Stephen Trussel. 2013. The Austronesian Comparative Dictionary: A Work in Progress. *Oceanic Linguistics* 52(2). 493–523. DOI: 10.1353/ol.2013.0016. <https://www.trussel2.com/ACD/>.
- Braun, Angelika, Nathalie Elsässer & Lea Willems. 2023. Disfluencies revisited – are they speaker-specific? *Languages* 8(3). 155. DOI: 10.3390/languages8030155.
- Candea, Maria, Ioana Vasilescu & Martine Adda-Decker. 2005. Inter- and intra-language acoustic analysis of autonomous fillers. In *Proceedings of disfluency in spontaneous speech workshop'05*, 47–52. Aix-en-Provence. <https://shs.hal.science/halshs-00321914>.
- Cheung, Lawrence Yam-Leung. 2015. Uttering the unutterable with *wh*-placeholders. *Journal of East Asian Linguistics* 24(3). 271–308. DOI: 10.1007/s10831-014-9130-x.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Corley, Martin & Oliver W. Stewart. 2008. Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Linguistics and Language Compass* 2(4). 589–602. DOI: 10.1111/j.1749-818X.2008.00068.x.
- Corminboeuf, Gilles & Laure Anne Johnsen. 2023. L’art de parler vague: Le N ‘truc’ comme compromis stratégique [the art of vague language: The N ‘truc’ as a strategic compromise]. In Dorothée Aquino-Weber, Sara Cotelli Kureth, Andres Kristol, Aurélie Reusser-Elzingre & Maguelone Sauzet (eds.), « *coum’on étèila que kòoule.... Come una stella cadente... Comme une étoile filante...* [Like a shooting star – A collection in the memory of Federica Diémoz] » *Mélanges à la mémoire de Federica Diémoz*, 235–262. Genève: Droz.

- Cutting, Joan (ed.). 2007. *Vague language explored*. Basingstoke: Palgrave Macmillan.
- Däritz, Chris Lasse. 2018. On *kim*, *kimne:-* and *kaña:-*. Three placeholder items in Dolgan. *Turkic Languages* 22. 286–302.
- Dimock, Laura. 2010. Fillers and placeholders in Nahavaq. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 119–138. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.06dim.
- Döhler, Christian. 2025. *That* placeholder in Komnzo. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 275–313. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697589.
- Egorova, Maria A., Elena Rudnitskaya & Elena Klyachko. 2021. *Code-switching vs borrowing: The case of hesitation markers in Evenki* (Paper presented at Discourse-Pragmatic Variation and Change 5). Melbourne.
- Enfield, N. J. 2003. The definition of WHAT-d'you-call-it: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Erker, Daniel & Lee-Ann Marie Vidal-Covas. 2022. What we say when we say nothing at all: Clues to contact-induced language change in Spanish conversational pause-fillers. *Estudios del Observatorio/Observatorio Studies* 80. 1–29. DOI: 10.15427/OR080-09/2022EN.
- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 1–9. Amsterdam: John Benjamins.
- Gaby, Alice R. 2017. *A Grammar of Kuuk Thaayorre* (Mouton Grammar Library 74). Berlin: De Gruyter Mouton.
- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Gerhalter, Katharina & Abir Salaaoui. 2020. Historia del arabismo fulano, -a [The history of the Arabism fulano, -a]. In Cristóbal José Álvarez López, María Carrillo Rivas, Diego Jiménez Palmero, María Méndez Oreense, Adriana Moratinos Flórez, María Soledad Padilla Herrada, Víctor Pérez Béjar, Marta Rodríguez Manzano, Ana María Romera Manzanares & Natalia Silva López (eds.), *Lingüística prospectiva: Tendencias actuales en estudios de la lengua entre jóvenes investigadores* [Prospective linguistics: Current trends in language studies among young researchers], 215–232. Sevilla: Editorial Universidad de Sevilla.

- Graziano, Maria & Marianne Gullberg. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology* 9(879). 1–17. DOI: 10.3389/fpsyg.2018.00879.
- Haude, Katharina. 2006. *A Grammar of Movima*. Nijmegen: Radboud University. (Doctoral thesis).
- Hayashi, Makoto. 2003. Language and the body as resources for collaborative action: A study of word searches in Japanese conversation. *Research on Language and Social Interaction* 36(2). 109–141. DOI: 10.1207/S15327973RLSI3602_2.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Heath, Jeffrey. 1999. *A grammar of Koyra Chiini : The Songhay of Timbuktu* (Mouton Grammar Library 19). Berlin: Mouton de Gruyter.
- Heath, Jeffrey. 2008. *A grammar of Jamsay* (Mouton Grammar Library 45). Berlin: Mouton de Gruyter.
- Heath, Jeffrey & Aminata Ouattara. 2021. *A grammar of Tiefo-D of Daramandugu: Niger-Congo language, Burkina Faso*. Ann Arbor: Language Description Heritage Library.
- Hennecke, Inga & Wiltrud Mihatsch. 2022. Between placeholder and filler: Degrees of prosodic prominence of French *machin* and *truc*. *Pragmatics & Cognition* 29(2). 297–323. DOI: 10.1075/pc.22003.hen.
- Huang, Huei-ju & Michael Tanangkingsing. 2005. Repair in verb-initial languages. *Language and Linguistics* 6(4). 575–597.
- Jacques, Guillaume. 2021. *A grammar of Japhug*. Berlin: Language Science Press.
- Jehoul, Annelies, Geert Brône & Kurt Feyaerts. 2016. Gaze patterns and fillers. Empirical data on the difference between Dutch ‘euh’ and ‘euhm’. In *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*. Copenhagen: LiU Electronic Press.
- Joseph, Brian D. 1997. On the linguistics of marginality: The centrality of the periphery. *Chicago Linguistic Society* 33. 197–213.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Khurshudyan, Victoria G. 2006. *Sredstva vyraženija xezitacii v ustnom Armjanskom diskurse v tipologičeskoj perspektive [Means of expressing hesitation in oral Armenian discourse in a typological perspective]*. Moscow: Russian State University for the Humanities. (Doctoral thesis).

- Khurshudyan, Victoria G. & Vera I. Podlesskaya. 2006. Armjanskoe *ban* kak diskursivnyj marker rečevogo sboja [Armenian *ban* as a discourse marker of speech disfluency]. *Revue arménienne humanitaire* 1. 21–42.
- King, Gareth. 2003. *Modern Welsh: A Comprehensive Grammar*. London: Routledge.
- Kirjavainen, Minna, Ludivine Crible & Kate Beeching. 2022. Can filled pauses be represented as linguistic items? Investigating the effect of exposure on the perception and production of *um*. *Language and Speech* 65(2). 263–289. DOI: 10.1177/00238309211011201.
- Kitano, Hiroaki. 1999. On interaction and grammar: Evidence from one use of the Japanese demonstrative *are* ('that'). *Pragmatics* 9(3). 383–400.
- Kluge, Angela. 2017. *A grammar of Papuan Malay* (Studies in Diversity Linguistics 11). Berlin: Language Science Press. DOI: 10.5281/zenodo.376415.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hödl & Thomas Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- Klyachko, Elena. 2025. Placeholders versus general extenders in Tungusic languages. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 41–74. Berlin: Language Science Press. DOI: ??.
- Kosmala, Loulou. 2024. *Beyond Disfluency: The interplay of speech, gesture, and interaction*. Amsterdam: John Benjamins. DOI: 10.1075/ais.11.
- Kosmala, Loulou & Ludivine Crible. 2022. The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1). 216–239. DOI: 10.1177/00238309211010862.
- Lahaussois, Aimée. to appear. Methods and tools for (meta)grammaticography. In David Bradley, Katarzyna Dziubalska-Kołaczyk, Camiel Hamans, Ik-Hwan Lee & Frieda Steurs (eds.), *Contemporary linguistics. Integrating languages, communities, and technologies*. Leiden: Brill.
- Larrivée, Pierre. 2009. What nominal phrases are all about: The atypical case of indefinite pronouns with a determiner. *Romanische Forschungen* 121(1). 3–19.
- LeSourd, Philip S. 2003. The noun substitute in Maliseet-Passamaquoddy. In Blair A. Rudes & David C. Costa (eds.), *Essays in Algonquian and Siouan linguistics in memory of Frank T. Siebert, Jr*, 141–163. Winnipeg: University of Manitoba Press.
- Lickley, Robin J. 2015. Fluency. In Melissa A. Redford (ed.), *The handbook of speech production*, 445–474. Malden: Wiley. DOI: 10.1002/9781118584156.ch20.

- Maclay, Howard & Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *WORD* 15(1). 19–44. DOI: 10.1080 / 00437956.1959.11659682.
- Maisak, Timur A. 2023. Leksičeskie “zamestiteli” i “rasprostraniteli” v rutul’skom jazyke: Materialy k vnutrigenetičeskoj tipologii [«Placeholders» and «general extenders» in Rutul: Towards an intragenetic typology]. In Irina Khomchenkova (ed.), *Malye jazyki v bol’soj lingvistike. Vypusk 5 [Minor languages in general linguistics. Issue 5]*, 134–144. Moscow: Buki Vedi.
- Matić, Dejan. 2008. A note on hesitative forms in two Ēven dialects. In O. A. Osipova, A. V. Didenko & Elena A. Krjukova (eds.), *Sravnitel’no-istoričeskoe i tipologičeskoe izučenie jazykov i kul’tur. Materialy meždunarodnoj naučnoj konferencii XXV Dul’zonovskie čtenija [Comparative-historical and typological studies of language and culture. Materials from the international scientific conference 25th Dul’zon Readings]*, 234–239. Tomsk: Tomskij gosudarstvennyj pedagogičeskij institut.
- Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.
- Mauri, Caterina & Andrea Sansò. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. *Folia Linguistica* 52. 1–35. DOI: 10.1515/flih-2018-0001.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- Mihatsch, Wiltrud. 2006. *Machin, truc, chose : La naissance de marqueurs pragmatiques [Machin, truc, chose: The birth of pragmatic markers]*. In Martina Drescher & Barbara Job (eds.), *Les marqueurs discursifs dans les langues romanes : Approches théoriques et méthodologiques*, 153–172. Frankfurt am Main: Peter Lang.
- Mihatsch, Wiltrud. 2024. Sustantivos comodines como recursos discursivos: El caso de *chisme* [Placeholder nouns as discourse devices: The case of *chisme*]. *Filología* 1(56). 103–126. DOI: 10.34096/filologia.n56.15082.
- Mithun, Marianne. 2025. Fillers and beyond. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 341–371. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697593.
- Nagaya, Naonori. 2022. Beyond questions: Non-interrogative uses of *ano* ‘what’ in Tagalog. *Journal of Pragmatics* 190. 91–109. DOI: 10.1016/j.pragma.2022.01.007.
- Navarretta, Costanza. 2016. The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations. In *Proceedings from the*

- 3rd European symposium on multimodal communication, Dublin, September 17–18, 55–61. Dublin: LiU Electronic Press.
- Overstreet, Maryann & George Yule. 2021. *General extenders: The forms and functions of a new linguistic category*. Cambridge: Cambridge University Press.
- Pakendorf, Brigitte. 2025. A corpus-based investigation of fillers in Negidal (Northern Tungusic). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 75–129. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697579.
- Palacios Martínez, Ignacio M. & Paloma Núñez Pertejo. 2015. “Go up to miss thingy”. “He’s probably like a whatsit or something”. Placeholders in focus. The differences in use between teenagers and adults in spoken English. *Pragmatics* 25. 425–451. DOI: 10.1075/prag.25.3.05pal.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Podlesskaya, Vera I. & Andrej A. Kibrik. 2009. Rečevye sboi i zatrudnenija [Speech disfluencies]. In Vera I. Podlesskaya & Andrej A. Kibrik (eds.), *Rasskazy o snovidenijax. Korpusnoe issledovanie ustnogo russkogo diskursa [Stories about dreams. A corpus investigation of Russian oral discourse]*, 177–218. Moscow: Jazyki slavjanskix kul'tur.
- Podlesskaya, Vera I. & Nikolay A. Korotaev. 2022. “Étu... Kak ejo... Perepisku Éngelsa s etim... Kak ego, djavola... S Kautskim” - A case study of a placeholder in Russian. *Russian Speech = Russkaja Reč'* 5. 64–78.
- Ponsonnet, Maïa. 2025. Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 315–340. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697591.
- Rice, Alexander. 2025. *Mashti*: A multipurpose filler in Northern Pastaza Kichwa. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 373–439. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697595.
- Rose, Françoise. 2025. One more thing ‘thing’ can do in Tupí-Guaraní languages: The Teko filler. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 441–484. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697597.
- Rubino, Carl. 1996. Morphological integrity in Ilocano: A corpus-based study of the production of polymorphemic words in a polysynthetic language. *Studies in Language* 20(3). 633–666. DOI: 10.1075/sl.20.3.05rub.

- Seraku, Tohru. 2020. Placeholders in Yoron Ryukyuan: A view from Functional Discourse Grammar. *Lingua* 245. 102852. DOI: 10.1016/j.lingua.2020.102852.
- Seraku, Tohru. 2022a. Interactional and rhetorical functions of placeholders: A relevance-theoretic approach. *Journal of Pragmatics* 187. 118–129. DOI: 10.1016/j.pragma.2021.10.023.
- Seraku, Tohru. 2022b. Referring to arbitrary entities with placeholders. *Pragmatics* 32(3). 426–451. DOI: 10.1075/prag.20076.ser.
- Seraku, Tohru. 2023. Creating placeholders on the fly: A metarepresentational approach. *Journal of Pragmatics* 210. 109–121. DOI: 10.1016/j.pragma.2023.03.021.
- Seraku, Tohru. 2024. Placeholders in crosslinguistic perspective: Abilities, preferences, and usage motives. *Linguistics* 62(6). 1473–1498. DOI: 10.1515/ling-2023-0068.
- Seraku, Tohru. 2025. Placeholders and interjective hesitators: A crosslinguistic and functional approach. *Studies in Language* Online First. DOI: 10.1075/sl.23057.ser.
- Sjånes Rødvand, Linn Iren. 2024. *Patani: Description of a South Halmahera–West New Guinea language*. Oslo: University of Oslo. (Doctoral thesis).
- Tárnýiková, Jarmila. 2019. English placeholders as manifestations of vague language: Their role in social interaction. *Brno studies in English* 45(2). 201–216.
- Vallejos Yopán, Rosa. 2023. From demonstrative to filler: *Este* in Amazonian Spanish and beyond. *Linguistics* 61(3). 651–678.
- van Driem, George. 1993. *A Grammar of Dumi*. Berlin: De Gruyter Mouton.
- Ventayol-Boada, Albert. 2025. Copula or placeholder: A qualitative and quantitative analysis of *λe*- in Kolyma Yukaghir discourse. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 131–168. Berlin: Language Science Press. DOI: ??.
- Visser, Eline. 2025. Form and frequency of Kalamang fillers. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 251–273. Berlin: Language Science Press. DOI: ??.
- Vogel, Petra M. 2020. *Dingsbums and Thingy*. Placeholders for names in German and other languages. In Lívia Körtvélyessy & Pavol Štekauer (eds.), *Complex words. Advances in morphology*, 362–383. Cambridge: Cambridge University Press.
- Watanabe, Honoré. 2010. Fillers and their relevance in describing Sliammon Salish. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 173–187. Amsterdam: John Benjamins.

- Williams, Nicholas. 2009. Toward a linguistic anthropological account of deixis in interaction: *Ini* and *itu* in Indonesian conversation. *Colorado Research in Linguistics* 22. 1–23.
- Wouk, Fay. 2005. The syntax of repair in Indonesian. *Discourse Studies* 7(2). 237–258. DOI: 10.1177/1461445605050368.
- Zhao, Yuan & Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. In *Proceedings of DiSS'05: Disfluency in spontaneous speech workshop*, 179–182. Aix-en-Provence: Université de Provence.

Chapter 2

Placeholders versus general extenders in Tungusic languages

Elena Klyachko

Higher School of Economics, Moscow

The paper describes two classes of discourse markers, namely placeholders and general extenders, in Tungusic languages (a family of endangered languages spoken in Russia and China). The aim of the paper is to provide an exhaustive list of stems, covering their etymology, mirroring features (i. e. copying morphology) and target stem restrictions. Although placeholders and general extenders both belong to the vague language category, they demonstrate different mirroring behaviour, which I will demonstrate. I argue that the use of the interrogative pronoun stem as a placeholder may have developed from the general extender use.

keywords: vague language, Northeast Asia, whatchamacallit, and such

1 Introduction

Placeholders (i. e. lexical forms that replace a target item) and general extenders (items used at the end of lists like “and so on” or “and so forth”, Overstreet & Yule 2021) are two categories of discourse markers, which have much in common. Both can be called *vague language* (see Tárnyiková 2019 for the term regarding placeholders) utilized to avoid naming a specific word. The reasons for the avoidance are manifold. The speaker may either not be able to recall a specific word, or they do not want to pronounce it, or they would like to refer to some open-ended list without specifying the exact word.

The interaction between placeholders and general extenders has already drawn the attention of linguists. Ganenkov et al. (2010) describe placeholders and general extenders in Udi and Agul (both belonging to the Northeast Caucasian



family). Their hypothesis is that in Udi and Agul, general extenders developed from placeholders, with the process being independent in both languages. The authors explain the similarity of the development by the common typological profile of Udi and Agul, which are both highly agglutinating, left-branching, and make great use of light verb constructions. It is therefore interesting if other languages with similar typological features exhibit similar patterns for placeholders and general extenders. The languages belonging to the Tungusic family are possible candidates for the research, being highly agglutinating and left-branching.

The following two Evenki (< Tungusic) examples demonstrate the parallels between placeholders and general extenders¹:

- (1) Evenki (Kazakevich, Klyachko, et al. 2023)

ə-wki: *minə* *e:-ra* *əyi-fko:n-ə*
NEG-PTCP.HAB 1SC.ACC what-PTCP.NEG play-CAUS-PTCP.NEG
'She did not **do** whatchamacallit, let me play.'

- (2) Evenki (Kazakevich, Klyachko, et al. 2023)

toyo-wo o:kin=da: *e:-wko:* *əru-t* *janyu-w-rə*
fire-ACC when=FOC NEG-PTCP.HAB bad-ADVZ talk-PASS-PTCP.NEG
ə-wko: *e:-w-rə=da:*
NEG-PTCP.HAB what-PASS-PTCP.NEG=FOC
'It should never be talked to fire badly or done similarly.'

In both (1) and (2), an *e:(kun)-* ('what') stem is used. In the first case, it substitutes the lexeme the speaker could not produce at the moment of speaking whereas in the second case, it refers to an open-ended list of actions which are all not allowed when communicating with fire. In both cases, we can observe *mirroring*. Podlesskaya (2010) uses this term with regard to copying inflectional and/or derivational affixes from the **target**, i. e. the constituent which is delayed or not pronounced at all. I think that the term can also be used for general extenders, when the general extender, used in an enumeration, copies the properties of the other members. However, the exact mirroring properties are different in examples (1) and (2). The wordform with the *e:-* 'what' stem is used in the negative construction in both examples, so it has the negative participle marker, which is

¹In this paper, I use **bold** for the placeholder or the general extender in examples. In case of placeholders, I use underline for the target word. In case of general extenders, I use underline for the first members of the enumeration (which the general extender often mirrors).

obligatory in this construction. At the same time, the voice marker is only found in the general extender example (2) but not in the placeholder one (1).²

Mirroring does not always have to be full, with Podlesskaya (2010) distinguishing between *partial* and *full mirroring*. Therefore, the question is whether the difference in the placeholder and general extender mirroring properties is a coincidence or a reflection of the underlying difference between placeholders and general extenders.

The aim of this paper is to describe the etymology, the grammatical features and functions of placeholders and general extenders in Tungusic languages as well as compare them. As a result, I would like to demonstrate the connections between these two categories, namely the derivation of one of the categories from the other one.

1.1 Tungusic languages

The Tungusic family is a family of endangered languages spoken nowadays over vast areas of Russia and China (Hölzl & Payne 2022). The details of the internal classification of Tungusic languages are still a matter of debate (see Oskolskaya et al. 2022 for a discussion). The sources used for this paper are summarized in Table 1. These are mostly oral corpora, grammars and sometimes papers, not specifically on placeholders but containing language examples. Unfortunately, the original sentence translations sometimes lack words like “whatchamacallit”. In such cases, I will use angle brackets to insert the translations for the placeholders. This paper considers nine Tungusic idioms. I do not take into account the Solon language as well as the dialects of Evenki and Nanai spoken in China due to the scarcity of data. Written idioms like Jurchen are not considered because little is known about spontaneous discourse.

1.2 Challenges in studying placeholders and general extenders

Studying placeholders and general extenders may be challenging due to several reasons. Firstly, placeholders, differently from general extenders, tend to be omitted in published texts and are often not mentioned in grammatical descriptions, having low prestige among both native speakers and linguists. It is also hard to perform elicitation experiments because in many cultures, including for instance Russian or Evenki cultures, placeholders seem to be a characteristic feature of “poor” speech. Speakers tend to avoid pronouncing placeholders when

²Anonymous reviewers also remark the repetition of the negative auxiliary in the general extender case. Unfortunately, there are too few examples of such repetitions to draw any conclusions of whether the general extender use favours such repetitions or not.

Table 1: Tungusic languages and the sources used

Language	ISO	Corpora	Grammatical descriptions
Evenki	evn	Kazakevich, Klyachko, et al. 2023 (mostly Northern and Southern dialect groups); Shakhovtsov et al. 2023 (Northern, Southern and Eastern dialect groups); Däbritz & Gusev 2021 (Northern and Southern dialect groups); elicitation	Klyachko (2022)
Negidal	neg	Pakendorf & Aralova 2017	Khasanova & Pevnov (2003)
Even	eve	Nikolaeva et al. (2019)	Matić (2008)
Udihe	ude	Nikolaeva et al. (2019)	Šneider (1936) Šneider (1937) Nikolaeva & Tolskaya (2001) Tolskaya & Tolskaya (2008)
Oroch	oac	Kazama (1996)	
Nanai	gld	Bel'dy & Bulgakova (2012)	Kazama (2007) Kazama (1993)
Kur-Urmi Nanai		Egorova 2017–2021	
Ulch	ulc	Stoyanova 2017–2021	Petrova (1936)
Ulta	oaa	Ikegami (2010), Yamada (2011), Yamada (2015), Yamada (2017), Gusev & Zhornik 2023, Kazakevich, Polivanov, et al. 2023	
Sibe	sjo	Kogura (2021); Kogura (2022)	Zikmundová (2013)

they control the discourse situation, and drawing attention to placeholder may make them upset. At the same time, the whole Tungusic language family is endangered, with the majority of Tungusic speakers being at least bilingual. This influences greatly the way hesitation markers are used. For example, it has been noticed that code-switching in bilingual speech is often accompanied by various hesitation and monitoring markers (see Hlavac 2011, where Croatian-Australian English bilingualism is studied). We can find such examples in the oral corpora, too. A brief discussion of such examples will be provided in 2.1.5.

Mirroring phenomena themselves are also challenging. Firstly, the target is not always pronounced during the recording or clarified after the recording. For instance, in the Evenki language corpus (Kazakevich, Klyachko, et al. 2023) the target was actually realized only in 60% of all cases for nouns and in 56% of all cases for verbs. Secondly, the investigation of mirroring is limited to the morphological features of the targets which occur in the corpora.

That said, I will still use mostly oral corpora for studying placeholders and general extenders. As I have already mentioned above, placeholders are often “discriminated” in literary texts, so using oral corpora is the best way to study them.

2 Tungusic placeholders

2.1 Tungusic placeholder stems

In Tungusic languages, several stems can be used as placeholders. All stems discussed below, except for the borrowed stems, are shown in Figure 1.

2.1.1 *aŋə/aŋi/anu*

This stem³ can be found in Evenki (3) (Klyachko 2022), Udihe (4) (Šneider 1936: 16, Nikolaeva & Tolskaya 2001: 361), Oroch (5), and Uilta (6) (Ozoliņa & Boldyrev 2001: 6):

- (3) Evenki (Kazakevich, Klyachko, et al. 2023)

<i>həwəki:</i>	<i>hələ</i>	<i>aŋi-l-duk</i>	<i>s'ita-l-duk</i>	<i>o-d'a-fki:</i>	<i>bi-s'</i> ə
God	INTJ	PH1-PL-ABL	clay-PL-ABL	make-IPFV-PTCP.HAB	be-PTCP.ANT
<i>bəjə-l-bə</i>					

human-PL-ACC

‘Well, God made people out of whatchamacallit, out of clay.’

³I will gloss it as PH1.

- (4) Udihe (Nikolaeva & Tolskaya 2001: 361)
ge oloktɔ:-ni belie aŋi-we zakta-wa
well cook.PST-3SG fairy PH1-ACC porridge-Acc
'Well, the fairy cooked this, what it's called, porridge.'
- (5) Oroch (Kazama 1996: text 12)
յԵհԵ-մ=ձԵՅ բա-խ-իս այի-վա օմօ կավա
jene-m=dee baa-xa-ji aŋi-wa ... omo kawa
go-CVB=FOC find-PST-SG PH1-ACC ... one house
'Walking, he found whatchamacallit... one house'.⁴
- (6) Ulta (Kazakevich, Polivanov, et al. 2023)
čōtči tar šof'or nari anu bāru-ni val
then that driver.R person PH1 in.direction-3SG.POSS Val
bāru-ni zvoni-ri-ni val gasa-tai-ni
in.direction-3SG.POSS phone.R-PRS-3SG Val settlement-ALL-3SG.POSS
'Then that driver phones in the direction of whatchamacallit, in the direction of Val, to the village of Val.'

The targets from the corpora include nouns ((3), (4), (5), (6)), verbs ((8), (9)) and, in rare cases, other parts of speech, such as adverbs (see Klyachko 2022). The frequency of this placeholder in the corpus of (Kazakevich, Klyachko, et al. 2023) is 340 per 37108 running words, which is about 9 times per thousand words. It is also worth noting that in Evenki, the use of the *aŋi* stem depends on the dialect, with its being rare or completely unknown in the Eastern Evenki dialects (Klyachko 2022). The corpus is not balanced regarding Eastern and non-Eastern data, with the non-Eastern data prevailing in it, and *aŋi*'s frequency for the non-Eastern dialect subcorpus is higher (about 10 per thousand words).

In the Ulta corpus (Gusev & Zhornik 2023), the frequency of *anu* is 75 per 18838 words, which is about 4 per thousand words. The lower frequency as compared to the Evenki corpus is due to the fact that some texts in the Ulta corpus are published narratives from Petrova (1967), which do not contain any placeholders, perhaps because of T. Petrova's editing the texts or due to the conditions of how

⁴Interestingly, this is the nominative form (*kawa(n)* 'cabin'), although the placeholder has the accusative suffix. However, the original audio recording shows that there was actually code-switching to Russian before pronouncing the word *kawa*, when the speaker was trying to recall the word.

she was recording them (they may have been dictated, which reduced the level of spontaneity). The target is overtly expressed in about 70% of the examples.

Unfortunately, I cannot measure the frequency of *aŋi* for Udihe and Oroch due to the lack of large open-access corpora.

When substituting nominal and verbal stems, *aŋa/aŋi/anu* can mirror the corresponding inflectional affixes. However, mirroring is not always full. Strikingly, the mirroring behaviour of *aŋa/aŋi/anu* seems to be uniform throughout the languages which have the stem. As regards nouns, *aŋa/aŋi/anu* stem usually mirrors case, number and possession, including indirect possession markers (sometimes called alienable possession markers, see Aralova & Pakendorf (2023) for an analysis). For example, in (7) the stem copies all possible nominal slots. In the first sentence, the speaker used the word *dēkso-ŋu-l-ba-ni* (otter-ALIEN-PL-ACC-3SG.POSS), which is what *anu-ŋu-l-ba-ni* may stand for in my opinion, as the other names of the animals are repeated. However, it is still a matter of doubt whether the speaker used the placeholder because of suddenly forgetting the word she used previously (a typical function of a placeholder) or because she simply wanted to refer to the target without specifying it.

(7) Uilta (Gusev & Toldova 2017–2021)

<i>no:ni duku-du-ni</i>	<i>oro: suli-ŋu-l-ba-ni</i>	
3SG house-LOC-3SG.POSS INTJ fox-ALIEN-PL-ACC-3SG.POSS		
<i>dēkso-ŋu-l-ba-ni</i>	<i>sə:pə-ŋu-l-ba-ni</i>	<i>ujlle:-tʃi</i>
otter-ALIEN-PL-ACC-3SG.POSS	sable-ALIEN-PL-ACC-3SG.POSS	work.PRS-3PL
<i>əlləutə mon-zi-tʃi</i>	<i><...> gə ilə:n-zi</i>	<i>ča</i>
very smoothen.PRS-3PL	<...> well three-INS	woman-COM that
<i>bəiŋə-ŋu-l-ba-ni</i>	<i>anu-ŋu-l-ba-ni</i>	
animal-ALIEN-PL-ACC-3SG.POSS	PH1-ALIEN-PL-ACC-3SG.POSS	
<i>sə:pə-ŋu-l-ba-ni</i>	<i>suli-ŋu-l-ba-ni</i>	<i>tʃipa:li</i>
sable-ALIEN-PL-ACC-3SG.POSS	fox-ALIEN-PL-ACC-3SG.POSS	all
<i>monzi-ya-tʃi</i>		
smoothen-PST-3PL		

'In his house, they work much, smoothen (the skins of) foxes, otters, sables. <...> With the three wives, they started to smoothen (the skin of) these animals, **and others**, sables, foxes.'

In verbal forms, the stem usually mirrors mood (8), tense, person, and number markers (9).

- (8) Evenki (Kazakevich, Klyachko, et al. 2023)
tʃajit tar tʃajit-pa tarə aŋi-wa:t tʃok-na:-ya:t
 bandit that bandit-ACC that.AC₁ PH1-IMPER.1PL.INCL kill-AND-PL.INCL
 'Let us do that thing, let us go and kill that bandit (=bear).'
- (9) Oroch (Kazama 1996)
mojoo=či маначаа aŋi-wa aŋi-xa-ňii loo-xo-ňii
 tojoo=tʃi mapatʃaa aŋi-wa aŋi-xa-ŋi loo-xo-ŋi
 then=FOC old.man PH1-ACC PH1-PST-3SG hang-PST-3SG
 'Then the old man did whatchamacallit to whatsisname, hung [fish to dry it].'

As we have already seen above (1), participial and converbial affixes can also be copied. Moreover, some other affixes, traditionally considered as derivational, are also copied, like the caritive marker in (10):

- (10) Uilta (Gusev & Toldova 2017–2021)
bi: mə:nə orkin-duk-ki un-zi-ni darama anu-lu ana
 1SG RFL bad-ABL-RFL say-PRS-3SG back PH1-CAR NEG
o-tči-mbi bəgʒi-lu ana
 become-PST-1SG leg-CAR NEG
 'Because of behaving badly I was left without whatchamacallit, without my legs.'

The question is which markers are **not** usually copied. The derivational markers that are mirrored are usually those which are both highly frequent and have little selection restrictions, i. e. combine with a great number of stems often belonging to different parts of speech, like the caritive marker, the adjectivizer as well as the diminutive or the intensifier markers. More specialized markers, such as verbalizers or nominalizers, are not usually copied. Interestingly, the voice marker, which is traditionally viewed as inflectional, is also not copied, at least in the examples from the corpora. For instance, the voice marker is not copied in the placeholder case (1), although it is copied in the general extender case (2). As regards particles, they are also sometimes copied as in (11), where the =dV: focus particle is used in the negation construction.

- (11) Evenki (Kazakevich, Klyachko, et al. 2023)
aŋi-n=da: ni= əmkə-n=də: as'in bi-so:-n
 PH1-3SG.POSS=FOC not.R.SLIP cradle-3SG.POSS=FOC NEG be-PST-3SG
 'There was no (his) whatchamacallit, (his) cradle.'

The etymology of the *aŋə/aŋi/anu* stem is enigmatic. For Evenki and Udihe, it has been reported as an interrogative pronoun (Bulatova & Grenoble 1999: 25) and an indefinite pronoun (Nikolaeva & Tolskaya 2001: 362). This would be a good etymology because interrogative and indefinite pronouns often become a source for placeholders (Podlesskaya 2010). However, if we look at the examples provided in the grammars, they are hardly “normal” indefinite or interrogative pronoun examples. On the contrary, they can rather be interpreted as typical placeholder examples. For instance, the authors remark that the following example (12) “can be translated differently according to the context”:

- (12) Evenki (Bulatova & Grenoble 1999: 25)

aŋi:-wa nujan bu:-rə-n
PH1-ACC 3SG give-NFUT-3SG

‘What/who/how many/what kind did he give?’

They also emphasize the fact that *aŋi* can also be used in assertive statements, “acquiring the role of a substantivized pronoun”. Moreover, Bulatova (2003) describes *aŋi*’s “more universal semantics” in comparison to the other Evenki interrogative pronouns. I think that the “semantic universality” of *aŋi*, its context-defined translations, and, importantly, its use in assertive statements prove that it is actually a placeholder and not just an ordinary interrogative pronoun.

As regards *aŋi* as an indefinite pronoun in Udihe, the situation is more difficult as the following Udihe example (13) unfortunately lacks context.

- (13) Udihe (Nikolaeva & Tolskaya 2001: 362)

aŋi-le ŋene:-mi
PH1-LOC walk.PST-1SG

‘I was walking somewhere.’

Nevertheless, I have not found any case of *aŋi* being used as an interrogative or an indefinite pronoun in the Evenki or Udihe text corpora. As Tolskaya & Tolskaya (2008) notice, *aŋi* “is abundant in spontaneous speech... but it does not occur in the autobiographical book by A. Kanchuga”, which would be odd for an ordinary indefinite pronoun. Next, the native speakers’ estimation of *aŋi* can also give a clue. Actually, many Evenki speakers’ reluctance to seeing *aŋi* in the transcription of their narratives is also typical of a low-prestige discourse marker but not of an ordinary interrogative pronoun. In addition, there exist standard interrogative and indefinite pronouns, which speakers use during elicitation experiments. Summing up, *aŋi*’s interrogative or indefinite pronoun meanings are

in my opinion just the result of how the authors of the grammars treated placeholder examples. It is also worth noting that the Russian placeholders used to translate the corresponding Tungusic ones are derived from pronominal stems. Such translations could also have influenced the linguists in analyzing *ayi* as a pronoun. The form *ayi* may have actually evolved from an interrogative or an indefinite pronoun but the stem is certainly not the standard interrogative or an indefinite pronoun in any Tungusic language at the synchronic level.

Typically, placeholder stems are etymologically connected with interrogative or demonstrative pronouns or abstract nouns meaning ‘thing’ (Podlesskaya 2010). Idiatov (2007) supposes that the Evenki *ayi* can be a word originally meaning ‘thing’ with the *-yi* formant, which is a possessive marker. The disadvantage of this hypothesis is the lack of the original ‘thing’ stem. I think that *ayi* can be also compared to the Amuric Nivkh *aŋ*, which is an interrogative pronoun used to refer to people and anthropomorphic creatures (Panfilov 1965: 253, Savyelyeva & Taksami 1970: 33). However, I should admit that this pronoun is not used as a placeholder in the corpus of oral texts in Amuric and Sakhalin Nivkh (Nakagawa 2023) in contrast with other interrogative pronouns, which are abundant there and used as placeholders. Still, pronoun borrowing between Nivkh and Tungusic languages is not impossible. For example, two Nivkh interrogatives may have been borrowed into Ulta (Hölzl 2018: 120). Therefore, it is theoretically possible that the Nivkh *aŋ* was initially borrowed as an interrogative pronoun but then turned into a placeholder. Placeholder borrowing itself is quite common, too, as I will show below.

2.1.2 *ujun/uŋ*

The form *ujun/uŋ*⁵ is another stem that can be found in some Evenki dialects (Klyachko 2022), in some Even dialects (Matić 2008, Robbek & Robbek 2005: 271), and in Negidal (Pakendorf 2025 [this volume]). In Even and Negidal, it is a general purpose placeholder ((14), (15)) for nominal or verbal stems, whereas in Evenki, it usually substitutes proper nouns like names of people or places ((16), see discussion in Klyachko (2022)).

- (14) Negidal (Pakendorf & Aralova 2017: AET_village_life 83)
- | | | | |
|---------------------------|--|---------------|-----------------|
| <i>tos a:cin bi-tca-n</i> | <i>o-ta-s</i> | <i>ujun-a</i> | <i>tosta-ja</i> |
| salt NEG | be-PST-3SG NEG-NEG.FUT-2SG PH2-NEG.CVB | salt-NEG.CVB | |
- ‘There was no salt, <so> you won’t do this, salt.’

⁵I will gloss it as PH2.

- (15) Even (Matić 2008)

min uŋ-il-bu [eń-til-bu] ... amaski badu-sn-i-tan
 1SG.OBL PH2-PL-1SG.POSS mother-PL-1SG.POSS ... back ride-LIM-PST-3PL
 'My **whatchamacallit** {my parents} ... went back.'

- (16) Evenki (Kazakevich, Klyachko, et al. 2023)

bajə uŋun-mə dəwit-pa ha:-Ø-ndə
 person PH2-ACC David-ACC know-NFUT-2SG
 'Friend, do you know **whatshisname**, David?'

Due to its proper name specialization in Evenki, *uŋun* is infrequent in the Evenki oral corpora. Being on the contrary the main placeholder in the Negidal language corpus, it occurs 496 times in a corpus of 41195 running words, that is about 12 times per one thousand words.

Just like *ayi*, the *uŋ(un)* wordform can copy inflectional and some frequent derivational affixes like the equative marker, as in (17).

- (17) Negidal (Pakendorf & Aralova 2017: APN_lechenie_ed 61)

eeee: net kip'ati-ja vari-ja kak uŋun-gatčin-mə ti kak
 INTJ NO.R boil.R-NFUT cook.R-NFUT like.R PH2-EQT-ACC like like.R
joktei-gatčin o:-knani-n
 tar-EQT become-CVB-3SG
 'No, one boils it, cooks it so that it becomes **like whatchamacallit**, like tar'

The etymology of the stem is once again enigmatic as it does not correspond to any known interrogative pronoun. The word can be found in the Evenki language dictionary as *ugun*⁶ (Vasilevich 1958: 625), with the following definition "who? what? a question on an unknown person or animal". Despite the wording of this definition, which may once again suggest an interrogative function of this word, there are actually no examples of its being used as an interrogative in a text, a grammar or a dictionary. I think that, similarly to the *ayə/ayi/anu* case, the Evenki *uŋun* is not an interrogative pronoun at the synchronic level, with the dictionary definition being misleading. In Cincius (1975–1977: 247), the Evenki *ugun*⁷ is compared to the Kur-Urmi Nanai *u:nəkə-*, an interrogative verb meaning 'what

⁶The interchangeability of intervocal *ŋ* and *g* is commonly attested in Evenki dialects.

⁷Cincius (1975–1977) must have used Vasilevich (1958: 625) as a source, which explains why the form is *ugun* and not *uŋun*.

to do?'. Unfortunately, no other parallels, like the Even or the Negidal one, are mentioned there, and it is still a question whether Tungusic languages have this stem. In this paper, I do not consider the Kur-Urmi Nanai *u:nəkə-* because, according to the dictionary and corpus data, it does not function as a placeholder. Still, the existence of this verb may suggest that *un(un)-* used to function as an interrogative pronoun. Its proper name specialization in Evenki must have developed later. However, a difficulty with the Kur-Urmi Nanai is its mixed-language nature (Hölzl & Payne 2022: 4) with a lot of Evenki influence. One option is that *un(un)-* was a question word in the common ancestor of Evenki, Even, Negidal and Nanai and then developed into a placeholder in the Ewenic branch comprising Evenki, Even, and Negidal. Another option is that *un(un)-* functioned as a question word in the common ancestor of Evenki, Even, and Negidal and was then borrowed into the Kur-Urmi Nanai from Evenki. In this latter case, it may have functioned as a question word in Evenki, too, and later developed into a placeholder in all the languages of the Ewenic branch, losing its question word properties.

2.1.3 Interrogative pronouns

Several Tungusic interrogative pronoun stems can be used as placeholders, namely: *xaj-*, *e:(kun)-* and *ja-*(see Hölzl 2018: 312–315 for a detailed description).

Etymologically, *xaj-* is reconstructed as **Kai* and is used as a placeholder in Nanai ((18), (19)) and Ulch (20). The stems *e:(kun)-* and *ja-* are reconstructed as **ja*. The form *e:(kun)-* functions as a placeholder in some dialects of Evenki ((21)), sometimes with the =*kana* or =*kana* particle (22)). The form *ja-* is a placeholder in some Even dialects (usually with the =*kana* particle, Matić 2008).

- (18) Nanai (Bel'dy & Bulgakova 2012: 216, 217)

эси=где тэй бэюн туў ту́муту-хэ-ни туў ту́муту-хэ-ни туў элэ
esi=gde tej bejun tuj tutu-xe-ni tuj tutu-xe-ni tuj ele
now=PTCL that so run-PST-3SG so run-PST-3SG so already
хай-ва-ни такто-ва-ни *ucu-xa*
xaj-wa-ni takto-wa-ni isi-xa
what-ACC-3SG.POSS barn-ACC-3SG.POSS reach-PST

‘Now that elk was running so fast, so fast, reached <her
whatchamacallit>, her barn’

- (19) Nanai (Bel'dy & Bulgakova 2012: 36, 37)

хай-у-го-и *киалако-у-го-и-ва*
xaj-ŋ-go-i *kialako-ŋ-go-i-wa*
 what-ALIEN-DESIG-1SG.POSS cropped.bone-ALIEN-DESIG-1SG.POSS-OBL
бүгү
bugu
give.back.IMP

'Give back the <whatsitsname>, the cropped bone!'

- (20) Ulch (Stoyanova 2017–2021)

na:n motor-wa xaj-ri-ni motor ŋən-i-wə-n ətəw-ri-ni
 3SG engine.R-ACC what-PRS-3SG engine.R go-PRS-ACC-3SG watch-PRS-3SG
 'He does whatchamacallit with the engine, watches how the engine is
 working.'

- (21) Evenki (Kazakevich, Klyachko, et al. 2023)

d'uktʃa e:kun-ma-n ham-na-∅ urkə-wə-n
 tent what-ACC-3SG.POSS close-NFUT-3PL door-ACC-3SG.POSS
ham-na-∅
 close-NFUT-3PL

'They close the tent's **whatsitsname**, its door.'

- (22) Evenki (Kazakevich, Klyachko, et al. 2023)

nu oro-r-duk bagar tari-ŋ-i nan aja-ma-t
 INTJ reindeer-PL-ABL perhaps that-ALIEN-ACC.RFL again good-INTS-ADVZ
e:-ŋna-∅-nni=kanan
what-HAB-NFUT-2SG=FOC

'Well, **you do** whatchamacallit <make a talisman> also very well from
 reindeer.'

There are very few examples of the corresponding 'what' stem used in a placeholder function in Negidal (23) and Udihe (24):

- (23) Negidal (Pakendorf & Aralova 2017: DIN_rite 28)

taj hujan e:kun-du-n səβəki-du-n təje-s ti:
 so forest what-DAT-POSS.3SG god-DAT-POSS.3SG treat-NFUT.2SG so
 'You treat <who>, the spirit of the taiga.

- (24) Udihe (Nikolaeva et al. 2019: Sisam Zauli and the hero)

a: *j'eu ute-bede ed'e-u kile diaŋ-ki-ni jeu=ke*
INTJ what this-how become.PF-3SG seagull say-PST-3SG what=INDEF
jewe-ni[uti uti] uti zoŋcula-ŋku o:-ni su:kte-zi
what-3SG this this shoot-NMLZ do.PST-3SG horsetail-INS
'The seagull said: «Ah, how did it happen? They⁸ <did whatchamacallit>, made an arrow out of a horse-tail⁹',

The frequency of *e:(kun)* as a placeholder is low in the Evenki text corpora. It is due to the fact that Southern and Northern dialect texts prevail in the corpora, whereas *e:(kun)-* is a typical Eastern placeholder. Still, *e:(kun)* does occur in non-Eastern texts but is far less frequent than *anji*.

Apart from the interrogative pronoun stems mentioned above, there is also *və* 'who' in Sibe, which is used as a placeholder for "a concrete proper name or the specific designation of a person" (Zikmundová 2013: 111), when the speaker has forgotten a name or does not want to pronounce it. For example, in (25), "the speaker is afraid of pronouncing a foreign name incorrectly".

- (25) Sibe (Zikmundová 2013: 111)

məzə-i və ... arien əm jašqaⁿ ji-y³i
1PL.INCL.GEN who ... Arienne one letter come-PERF
'<whatchamacallit> Arienne wrote me a letter'

2.1.4 Demonstrative pronoun placeholders

Demonstrative pronouns are typologically frequent sources for placeholders. For instance, in Russian, one of the most frequently used placeholders is *əmom* (*etot* 'this'), which has been borrowed into Tungusic languages, too (see §2.1.5). However, among the Tungusic languages covered in this survey, it is only Sibe that makes great use of demonstratives (*ək*, *səkei/skəi*) as placeholders, according to the detailed description in Zikmundová (2013), with (26) as one of the placeholder examples:

⁸The original clause is impersonal, this is why 3SG verbal markers are used.

⁹Despite its form, *jewe-* is probably not an accusative form, see Hölzl (2018: 324) for a discussion.

- (26) Sibe (Zikmundová 2013: 106)

ərχaʂəč da nan²-i favən-t dØš-kəŋ arq-aq^ü da sk²i
 this boy PTCL person-GEN law-DAT enter-PERF method-NEG PTCL PH
xova²-i diØrgit qotoN zə-m da dØš-k²i
 coffin-GEN inside ONOM say-CVB PTCL enter-PERF

‘The boy had no other choice but to follow the (other person’s) order, and so he entered **whatchamacallit** coffin with a thud.’

Zikmundová (2013) argues that Sibe placeholders can be used not only when the speaker has forgotten a word but also in case they do not want to pronounce it. For example, in (26) the word *xova²-i* ‘coffin-GEN’ was disturbing for the speaker. This avoidance function is typical of placeholders, see Cheung (2015), Seraku (2024).

2.1.5 Borrowed placeholders

Placeholder borrowing is not a rare phenomenon. For example, in Visser (2025 [this volume]), placeholder borrowing in Kalamang is discussed.

Tungusic languages demonstrate some borrowed placeholders, such as *eto* ('this' < Russian (borrowed into many languages in contact with Russian, cf. Ventayol-Boada 2025 [this volume] on Kolyma Yukaghir); ((27), (28), (29)) or *bolla* / *bo:la* (<*buolla*, which is a modal particle in Sakha, ((30), (31)).

These placeholders differ greatly from the “native” ones because of the lack of mirroring. In Sakha, *buolla* is a non-inflecting modal particle derived from a verbal stem. In Russian, *eto* can be a non-mirroring filler but mirroring is also possible with the same stem (*etot*). On the contrary, in the Evenki texts, *eto* is never inflected. Not surprisingly, it is sometimes hard to say if the borrowing is used as placeholder or as a discourse marker from another category. For example, in (27), it is used at the beginning of the phrase as a hesitative marker. In (28), *eto* is used together with the Russian borrowing *satun* ‘travelling bear’, so the question is whether the use of these two Russian words is an example of borrowing or code-switching. In (29), it is perhaps a placeholder. Actually, in most examples from the corpus of (Kazakevich, Klyachko, et al. 2023), *eto* is rather a hesitative at the beginning of the phrase or is used together with other Russian borrowings. Therefore, the lack of mirroring may be due to the fact that the word is a borrowing and not fully morphologically adapted or because it is actually not a borrowing but rather an example of code-switching.

- (27) Evenki (Kazakevich, Klyachko, et al. 2023)
eto jakutskij-la: əmə-ksək t'ot'a-m guni-pki: a
 this.R Yakutsk-ALL come-CVB aunt.R-POSS.1SG say-PTCP.HAB INTJ
ajnu:pki:
 ask-PTCP.HAB
 ‘Uhm, when I came to Yakutsk, my aunt said, oh, asked’
- (28) Evenki (Kazakevich, Klyachko, et al. 2023)
umno bi-tʃo:-n e:kun=ka nu hawal-d'i-a-ri: eto
 once be-PST-3SG what=FOC INTJ.R work.travel-IPFV-PTCP this.R
fatun
 travelling.bear.R
 ‘Well, once there was **whatchamacallit**, a travelling bear’
- (29) Evenki (Kazakevich, Klyachko, et al. 2023)
tadu: eto iʃə-t-tʃə-m=də:
 there this.R see-DUR-FUT-1SG=FOC
 ‘There, I will do **whatchamacallit**, I will see.’

As regards *bolla*, it is used widely in the Evenki dialects of the Far East and in the dialects of Even. A native speaker of Evenki¹⁰ even directly interpreted it as the borrowed Sakha word for the Evenki *aji*. However, it is also hard to say whether it is a placeholder or a hesitation filler in examples like (30) or the Lower Kolyma Even example (31), where the Sakha borrowing *bolla* is used together with the “native” interrogative placeholder:

- (30) Evenki (Kazakevich, Klyachko, et al. 2023)
a lamus o-ra-n bo:la
 INTJ deep.snow become-NFUT-3SG PTCL.SAH
 ‘Oh, it was deep snow, **whatchamacallit**’
- (31) Even (Sharina & Kuzmina 2018: 139)
ə tar ja-vra-ra-m bolla ikə-vrə-rə-m
 INTJ that what-HAB-NFUT-1SG PTCL.SAH sing-HAB-NFUT-1SG
 ‘Oh, <**I do whatchamacallit**>, I sing’

¹⁰V. Sabrski, the settlement of Tugur, p. c.

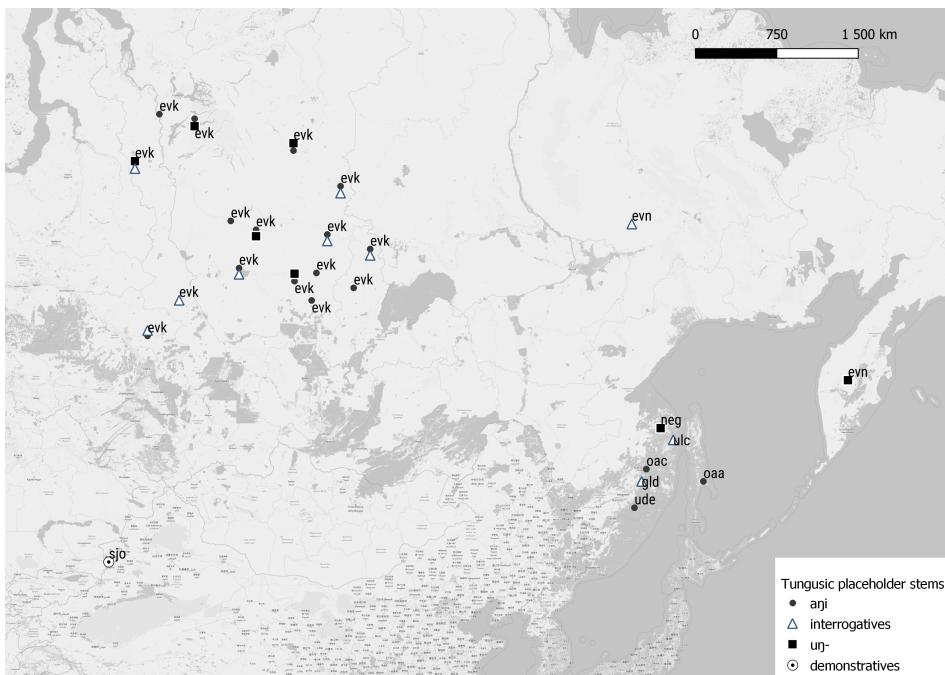


Figure 1: Placeholder stems in Tungusic languages

2.2 Tungusic placeholder functions and formal features

Judging from the corpora examples, the functions of the Tungusic placeholder stems are quite typical of placeholders in general. Tungusic placeholders are used in the following situations:

1. The speaker cannot recall a word but wants to preserve fluency and keep the floor (most examples like (3)).
2. The speaker does not want to pronounce a specific word due to extralinguistic reasons (26).
3. The speaker wants to draw the listener's attention due to some extralinguistic factors. For example, the speaker may want the listener to help them recall the target (like (16), see discussion in Klyachko (2022: 216)). The key point is the interaction between the speaker and the listener. In a way, it is an opposite strategy, as compared to 1. In 1, the speaker wants the listener to ignore the disfluency and just reconstruct the possible target. In 3, the speaker wants the listener to actively help them recall the

missing word. In Sibe (Zikmundová 2013: 106, 111), placeholders are sometimes used by speakers to draw the listener's attention to a new participant of the situation or to emphasize the target because it is important (32) or somewhat extraordinary, e. g. it is a foreign word.

- (32) Sibe (Zikmundová 2013: 106)

dučiurə-t sk²i tər ək miao bi vaq na
fourth.banner-DAT PH that PH temple be NEG INTER

'In the Fourth Banner (a settlement in the Qapqal Xibe Autonomous County) there is (that one, you know what) temple'

It is not always possible to derive the actual pragmatic situation from a corpus example. Still, dialogues from the corpus of Kazakevich, Klyachko, et al. (2023) provide much insight into the speakers' interaction and show that speakers actively use placeholders to draw the listeners' attention and ask them for help.

Despite the seeming abundance of various placeholder stems across the Tungusic family, all the stems, apart from the borrowed ones, behave in a very similar way. They can all substitute both nominal and verbal stems, except for the Evenki *ujun*, which has a specialization for proper nouns. As regards nouns, placeholders usually mirror their case, number, and possession markers. In verbal forms, they usually mirror mood, tense, person, and number markers but not the voice marker. Interestingly, they can also copy frequent derivational markers, such as the equative marker, as in (17). For instance, in the Evenki language corpus (Kazakevich, Klyachko, et al. 2023), nominal inflectional morphology is fully mirrored in about 76% of all cases where it is possible to estimate mirroring at all (i. e. where the target is present), whereas verbal inflectional morphology is fully mirrored in about 67% of the cases. There are only several sporadic examples of derivational morphology copying, although this may be explained by the comparative rarity of these markers.

The only language, which clearly has a separate placeholder with a specialized function (namely, *ujun* used specifically for proper nouns) is Evenki. In other cases, there is usually one most frequent placeholder for a dialect. For example, *aji* is used in some Northern and Southern Evenki dialects whereas *e:(kun)* is characteristic of most Eastern Evenki dialects.

3 Tungusic general extenders

3.1 Tungusic general extender stems

General extenders is originally a term suggested by M. Overstreet for English phrases like “and stuff like that” (Overstreet 1999). In Mauri & Sansò (2018), they are defined more precisely as elements located at the end of a non-exhaustive list, “whose meaning is indexical with respect to some underlying property P”. In most examples from this section, the original “list” usually comprises only one element but there may be more elements in it, like in (34).

Placeholder and general extender functions can sometimes be found for one stem (see Rose 2025 [this volume] for an analysis of *bae* in Teko). General extenders derived from the interrogative pronoun stem are widely used for both nominal and verbal phrases in the Tungusic languages. In all Tungusic languages, the interrogative pronoun stem can also form an interrogative verb ‘what to do’, so the verbal morphology is actually quite common for the stem. I will look at the mirroring properties of Tungusic general extenders and compare them with those of the placeholders.

In the Evenki example (33), the interrogative pronoun *e:(kun)* is used twice, once in its full form (with the *-kun* suffix) as a placeholder, and once in its short form as a general extender¹¹. As a general extender, it is used after *oldo-ŋi-l-wa* (fish-ALIEN-PL-ACC) to denote other goods, apart from fish, which were transported to the settlement of Valyok. Interestingly, it mirrors the number and case but not the indirect possession suffix. (34) is one of the rare examples where the enumeration comprises more than one member, with the members of the enumeration all referring to various items a well-to-do person was once thought to possess. The general extender copies the adjectivizer suffix here. In (35), the same stem is used to refer to the skills that people usually learn at school, such as reading or writing. The members of the enumeration are thus verbal phrases.

- (33) Evenki (Kazakevich, Klyachko, et al. 2023)
- | | | |
|---|---------------------|----------------|
| <i>walok-tula: tozə e:ku-r-wa</i> | <i>oldo-ŋi-l-wa</i> | <i>e:-l-wa</i> |
| Valyok-ALL also what-PL-ACC fish-ALIEN-PL-ACC what-PL-ACC | | |
| <i>əmə-wu-pki:-l</i> | <i>bi-ʈʃo-l</i> | |
| bring-TR-PTCP.HAB-PL be-PTCP.ANT-PL | | |
- ‘They used to bring whatchamacallit, fish and so on to Valyok’

¹¹Actually, full-form general extenders can also be found in the corpus, so it is a question whether this short and full-form distinction is coincidental or not.

- (34) Evenki (Kazakevich, Klyachko, et al. 2023)

nu oro-sⁱ o-sa d'u:-sⁱ e:-sⁱ
 INTJ.R reindeer-ADJ become-PTCP.ANT house-ADJ what-ADJ
o-sa

become-PTCP.ANT

‘Well, he became (a person) with reindeer, with a house and stuff.’

- (35) Evenki (Kazakevich, Klyachko, et al. 2023)

muldi:-ka-r ərəgərit e:-wa=da doku-d^ja-mi:=da
 cannot.do-NMLZ-PL at.all what-ACC=FOC write-IPFV-CVB
e:-d^ja-mi:=da
 what-IPFV-CVB

‘They could not do anything, could not write or do similar things’

The same *e:(kun)* ‘what’ stem is used as a general extender in Negidal ((36), (37)). In (36), it refers to the objects that can be used to produce light, like a torch or matches. It does not only mirror the nouns’s possessive marker but is also accompanied with the *=dV* focus particle. *=dV* is, on the one hand, typically used in the coordinating construction in Negidal and, on the other hand, marks negative polarity.

- (36) Negidal (Pakendorf & Aralova 2017: GIK_2tatarskoe 39)

man-mi fonariki-β=de e:kun-mi=da a:tcin
 RFL-1SG.POSS torch.R-1SG.POSS what-1SG.POSS=FOC NEG
 ‘I myself didn’t have a torch or anything.’

In (37), the same ‘what’-stem is used as a verb referring to the actions people usually perform when going to the forest, like picking berries. It mirrors the con-verb affix but not the *-lV* verbalizer, which derives ‘gather.berries’ from ‘berry’.

- (37) Negidal (Pakendorf & Aralova 2017: GIK_2tatarskoe 95)

taj təβ-lə-ja:n e:-ja:n əmə-dgi-ja-βun pan
 that like.this berry-VBLZ-CVB what-CVB come-REP-NFUT-1PL.EXCL also
siksə
 in.the.evening
 ‘We picked and did something, in the evening we came back’

In Even, *ja* ‘what’ is an interrogative which is also used as a general extender. In the Even example ((38)), the ‘what’-stem refers to the various ways of ornamenting bedclothes. Interestingly, it does not copy the participial *-ti* but copies

the nominal proprietive affix (*-lkan*). In (39), the verb with the ‘what’-stem refers to the actions connected with taking care of a reindeer. Interestingly, in (39), the proprietive suffix is mirrored but not the voice marker (see the discussion of voice marker mirroring in examples (1) and (2)).

- (38) Even (Nikolaeva et al. 2019: Glove and love)

<i>ere-k</i>	<i>te:den=de</i>	<i>butun-ni</i>	<i>[ońa:]</i>	<i>ońa:-tj-lkan</i>
this-NMLZ	bedclothes=FOC	all.SAH-POSS.3SG	paint	paint-PTCP-PROPR
<i>ia-lkan</i>				what-PROPR

‘And the bed was all decorated with patterns and stuff’

- (39) Even (Pakendorf et al. 2010: Top09_GNM_4_1.562)

<i>nøyman</i>	<i>oŋk-u-t-nikan</i>	<i>ia-nikan</i>
3SG.ACC	graze-TR-RES-CVB	what-CVB

‘I feed him et cetera.’

General extender constructions can also be found in Udihe. They make use of the interrogative *je* ‘what’. In (40), the general extender refers to the products of hunting like meat, blood or marrow. In (41), the ‘what’-stem verb denotes the actions which are usually associated with taking care of a baby, like rocking it. Unlike the previous examples from this section, where the ‘what’-word directly succeeds the enumeration, in (41), there is a coordinating construction: “rocked – crying, (general extender) – crying”, which also resembles (2). The repeating of *sojoi* resembles “recycling”, or repeating material, described by Podlesskaya (2010) for placeholders.

- (40) Udihe (Nikolaeva et al. 2019: Zabdala, an extraordinary snake)

<i>uta-zi</i>	<i>wa-i</i>	<i>diga-i</i>	<i>uta-zi</i>	<i>diga-Ø-li</i>	<i>ule:-we</i>	<i>je-we</i>
that-INS	kill-PTCP	eat-PTCP	that-INS	eat-3SG	meat-ACC	what-ACC
<i>diga-i</i>						eat-PTCP

‘Then they would eat what it had killed, the meat and all the rest.’

- (41) Udihe (Nikolaeva et al. 2019: Zabdala, an extraordinary snake)

<i>emu-si:-Ø-ni</i>	<i>sojoi-i</i>	<i>i:-si-Ø-ni=de</i>	<i>sojoi-i</i>
rock-IPFV-NFUT-3SG	cry-PTCP	what-IPFV-NFUT-3SG	cry-PTCP

‘The old man rocked him and did everything, but the child was still crying.’

Due to the small amount of available Oroch data, I could only find a nominal example (42), where the ‘what’-stem refers to the kinds of traces one could see. The focus particle copying resembles the Negidal example (36), where once again =dV can be interpreted as a coordinating particle or as a negative polarity particle. Notice that in (42), the hesitative *anj* is used too, contrasting with *jeu*. The former is used to fill in the pause, whereas the latter extends the enumeration started with the word *xokto* ‘footprint’.

- (42) Oroch (Kazama 1996: text 10)

хокто=даа=мааки **јэу=дээ** **аји**
xokto=daa=maaki jeu=dee anji
way=FOC=PTCL what=FOC HES
‘(There are) no footprints or anything <, well>’.

In Uilta, the corresponding interrogative pronoun is *xai* ‘what’. In (43), the ‘what’-wordform refers to various kinds of food. Notice the structural parallelism of this example with (41): in (43), the verb *dəp-tfi-pu* ‘we eat’ repeats after the general extender.

- (43) Uilta (Gusev & Toldova 2017–2021)

musi-l-ba **dəp-tfi-pu** **xai-l-ba** **dəp-tfi-pu** **un-zi-ni**
musi-PL-ACC eat-PST-1PL.EXCL what-PL-ACC eat-PST-1PL.EXCL say-PRS-3SG
‘We ate musi (kind of jelly made from skin fish), we ate **other things**, –
he said.’

In (44), the general extender refers to various parts of the object, mirroring the morphology on *do*-: ‘inner part’ . In (45), *xai-* refers to the actions which a grown-up woman traditionally performed, such as sewing.

- (44) Uilta (Gusev & Toldova 2017–2021)

təto:-ni *do*:-kke:-ni **xai-kke:-ni=da** *tſipal*
clothes-3SG inner.side-PROL-POSS.3SG what-PROL-POSS.3SG=FOC whole
bara: *bara:* *i:-xə-tfi*
very.much very.much enter-PST-3PL

‘They entered into her clothes and all over <lit.: into the inner part and other parts of her clothes>’

- (45) Uilta (Gusev & Toldova 2017–2021)

uže *nari* *o-tči-ni* *ulp*-i-ni **xaj-ri-ni**
already.R person become-PST-3SG sew-PRS-3SG what-PRS-3SG
‘She already grew up, sewed and did other things.’

In Nanai, the cognate *xaj* stem is used. In (46), the general extender refers to the equipment one wears when going to the forest, such as skis. In (47), the general extender refers to the set of actions a young man was traditionally thought to be able to do, such as shooting.

- (46) Nanai (Bel'dy & Bulgakova 2012: 36–37)

<i>эди-и</i>	<i>пулэ</i>	<i>сокта-ва-ни</i>	<i>эди</i>	<i>пулэ</i>
<i>edi-i</i>	<i>pule</i>	<i>sokta-wa-ni</i>	<i>edi</i>	<i>pule</i>
husband-POSS.RFL	extra	ski-ACC-POSS.3SG	husband	extra
<i>xaj-ва-ни</i>	<i>тэтүгү-хэ-ни</i>	<i>тавајки</i>	<i>дүйси</i>	<i>то-ха-ни</i>
<i>xaj-wa-ni</i>	<i>tetugu-xe-ni</i>	<i>tawanki</i>	<i>dujsi</i>	<i>to-xa-ni</i>
what-ACC-POSS.3SG	put.on-PST-3SG	afterwards	to.forest	go-PST-3SG
'She put on her husband's extra <u>skis</u> , her husband's extra stuff , and she went to the forest'				

- (47) Nanai (Kazama 1993: text 7))

<i>эй</i>	<i>пиктэ-н(и)=тэнни</i>	<i>эси=тэнни</i>	<i>лэкээ-чи=ү</i>	<i>хай-ри=ү</i>
<i>ej</i>	<i>pikte-n(i)=tenii</i>	<i>esi=tenii</i>	<i>lekee-tʃi-i</i>	<i>xaj-ri=i</i>
INTJ	child-POSS.3SG=PTCL	NOW=PTCL	arrow-VBLZ-PTCP	what-PTCP=PTCL
<i>ма-ло-ха-ни</i>				
<i>ta-lo-xa-ni</i>				
do-INCH-PST-3SG				
'The child started <u>to shoot and do other things like that</u> '				

Unfortunately, due to the small amount of Ulch data I could not find a clear general extender example from Ulch, apart from (53), which I analyze in more detail below.

The general extenders discussed in this section exhibit mirroring. They usually mirror inflectional morphology, often including verbal voice, as in (2) (but not in (39)). They sometimes mirror derivational morphology as well, as in (51). However, we can see that full mirroring does not always take place, for example, in (39), where the general extender refers to a set of actions akin to feeding. Naturally, these actions do not necessarily follow the causative model of 'graze' > 'feed'. This is why the transitive marker used in the 'feed' wordform is not copied. In (53), the general extender has the accusative marker lacking in the first member of the enumeration, which raises a question of how direct object marking works in case of general extenders. Unfortunately, at the time, there is too little data to study it.

Still, I suppose that general extender mirroring is generally fuller than placeholder mirroring. However, this should be proved using larger corpora for each language.

3.2 Borrowed general extenders

Borrowed general extenders seem to be harder to spot in the corpora. In Russian, there are multiword expressions like *i tak dalee* ('and so further') used as general extenders but they are rare in the Tungusic oral texts. In the Sakha language dictionary (Sakhatyla 2023), *эгин* (*eγin*), an adjective meaning 'different', is also defined as a "particle generalizing similar objects or actions appended to the main object or action". In this latter meaning, it is often written as *игин* (*igin*). This word can actually mirror the morphological features of the elements of the list as in (48).¹²:

- (48) Sakha (S. Makarov, p. c.)

<i>ommyy-бум</i>	<i>cup</i>	<i>астыы-быт</i>	игин-нии-бит
<i>ottuu-but</i>	<i>sir</i>	<i>astii-bit</i>	<i>igin-nii-bit</i>
mow-PRS.1PL	earth	obtain-PRS.1PL	etc.SAH-vBLZ-PRS.1PL
'We <u>mow grass, gather berries <lit. obtain earth food></u> etc.'			

igin occurs several times in the Even language corpus, as in (49):

- (49) Even (Pakendorf et al. 2010: Krivoshapkin_SP_oxota 019)

<i>ibga</i>	<i>ŋina-lkan</i>	igin	<i>bi-wre-n</i>	<i>ta-la</i>
good	dog-PROPR	etc.SAH	be-HAB-3SG	that-LOC
'He used to have a good dog <etc> there'				

Just like the borrowed placeholders, the borrowed general extender in Even does not mirror morphological features.

3.3 Tungusic general extender functions

The majority of the examples show that the main function of general extender is to refer to an open-ended list of objects and activities, which the speaker does not specify because they are a part of the background knowledge, as shown with the examples in the preceding section.

¹²This example was provided by S. Makarov.

However, general extenders have a number of other functions, too. One of these “extra” functions is the approximative one. In (50), *ia-* has an approximative function, because the speaker is not sure about the number of the days or perhaps it is actually not a strict rule of the rite. In (51), the function of *xai* is also approximative because the speaker is not sure if the oriole’s feathers can be called a “crest” in Udihe and is looking for a right word.

The question is whether these approximative usages of the ‘what’-stem are examples of placeholder use, or general extender use, or just of interrogative pronouns used as indefinites. I argue that the approximative constructions are derived from the general extender construction. Firstly, they are different from the normal placeholders formally. For example, in Even, the *=kanan* particle must accompany the ‘what’-placeholder. In Udihe, the ordinary placeholder is *ayi* and not *je*. Secondly, developing approximative function is common for general extenders (cf. Kim 2020). Finally, according to the grammars of Even (Malchukov 1995: 12), Negidal (Cincius 1982: 22) or Udihe (Nikolaeva & Tolskaya 2001: 353), interrogative pronouns are not generally used as indefinites without any other markers, requiring a particle like *=DV*.

Another construction, which is in my opinion connected with general extenders is the alternative question construction, analyzed by Tolskaya & Tolskaya (2008). In this specific case, the speaker uses ‘what’ to denote an alternative to the main action.

- (52) Udihe (ex:klyachko: 13 from Tolskaya & Tolskaya (2008))
guli-zeñe-fi=es ja-zanya-fi=es
leave-FUT-1PL.INCL=DIS what-FUT-1PL.INCL=DIS
‘Shall we leave or what?’

Overstreet (2020) calls English phrases like “or something” disjunctive general extenders, because they refer to a set of alternatives. Although she does not specifically cover alternative questions, I think that such questions are actually close to non-interrogative usages like (53), where *xaj-* stands for an alternative.

- (53) Ulch (Stoynova 2017–2021)
tımana=gdal um du:sə giwu=nü xaj-wa=nü wa:-ra
tomorrow=PTCL one tiger roe=INTER what-ACC=INTER kill-CVB
gažu-xa-ni
bring-PST-3SG
‘In the morning, one tiger brought either a roe or something.’

To sum up, in my opinion, there are three types of situations, which can all be classified as general extender constructions. In the first one, general extenders are used in the classical way at the end of an enumeration to refer to a non-exhaustive list of objects or actions. In the second one (“approximatives”) they refer to a non-exhaustive list of situations (like “nine days or something”, which can actually be unfolded as a list ‘nine days or eight days or ten days...’). In the third one (“alternative questions”) they refer to an non-exhaustive list of alternatives to the main situation. All these cases use the same ‘what’-pronouns, which contrasts with specialized placeholder stems in Evenki, Negidal, Even, Udihe, Oroch, and Uilta.

4 Comparing placeholders and general extenders

Tungusic placeholders and general extenders have somewhat similar functions, being used as “vague language”, when the speaker cannot or does not want to specify the exact target. They sometimes have the same stems and demonstrate mirroring behaviour. However, there are also differences between them.

Firstly, Tungusic general extenders are uniformly derived from the ‘what’-stem. On the contrary, placeholders have various stems in various Tungusic languages, with their etymology often not clear, sometimes suggesting borrowings. Secondly, their mirroring properties are different: general extenders tend to perform full mirroring, including the derivational morphology, whereas for placeholders, the mirroring is often partial. What is the reason for these differences? I think that the main explanation is just the fact that in case of general extenders, the first member(s) of an enumeration have already been pronounced. Therefore, copying is more mechanical. Furthermore, in case of placeholders, the target that is realized may be not the word the speaker was originally trying to recall.

As regards the stem uniformity vs. diversity, the explanation can be the fact that the use of ‘what’-stems for placeholders is influenced by the general extender use with ‘what’-stems drifting to the placeholder area and becoming the default vague language markers. This is opposite to the Udi and Agul case discussed in Ganenkov et al. (2010). The sporadic examples from Negidal (23) and Udihe (24) show that ‘what’-stems can actually be used as placeholders. However, due to the moribund status of both languages and the scarcity of the language data, we cannot see the possible development of the process. An anonymous reviewer suggested another possible explanation, i. e. an independent development of general extenders and placeholders from the interrogative pronouns, which are also known to function as indefinites in the Tungusic languages. However, I think that this is less probable because in many Tungusic languages the indefinite pronouns are not just bare interrogative stems. On the contrary, they usually require additional particles, lacking in the placeholder or general extender examples shown above. Another possibility is an independent development of placeholders and general extenders from an interrogative stem but the aforementioned sporadic examples are in favour of the first hypothesis.

Was there a dedicated Tungusic placeholder stem? Word forms like *ayi-* and *ujun-* are good candidates since they can be found in several Tungusic languages. On the other hand, the data analysis shows that placeholders are quite often being borrowed. Therefore, we cannot exclude the fact that, as shown above, *ayi-* was simply borrowed from Nivkh into neighbouring Tungusic languages in the Amur area.

5 Conclusions

The paper shows that Tungusic placeholders and general extenders may be thought of as two very similar classes of discourse expressions, both dealing

with uncertainty. However, they cannot be considered a continuous class. Firstly, placeholders exhibit much more diversity regarding the stems whereas general extenders are uniform. One possible explanation for the fact is that the original general extenders became placeholders in Nanai, Ulch and Uilta as well as some Even and Evenki dialects. The same process may be seen in several Udihe and Negidal examples but the languages are unfortunately moribund, with no grammaticalization (or pragmaticalization) process going on actively. Another peculiarity is the fact that placeholders seem to be much more prone to borrowing than general extenders. Specifically, there are borrowed Russian placeholders and Sakha particles in Evenki and Even. Moreover, *anj*, a placeholder stem with no good etymology, may be a Nivkh borrowing. Placeholder borrowing in modern texts can be explained by the fact that in case of attrition placeholders help maintain fluency, whereas code-switching to Russian and Sakha is also frequent in case of attrition.

Another difference between placeholders and general extenders is their mirroring behaviour. General extenders mirror more features, including e. g. verbal voice. An obvious explanation is that in case of general extenders, the speaker has already produced the wordforms in the enumeration, so it is easier for them to copy the affixes. As regards the placeholders, the speakers find themselves in a more difficult situation with the morphological features of the target not yet produced. The left-branching nature of the Tungusic languages must be crucial for placeholders: as Ganenkov et al. (2010) put it, “in left-branching languages syntactic dependents can appear before the processing difficulties in the head’s nomination occur”. Still, partial mirroring means that the markers, which are not determined by the syntactic dependencies only, may be difficult to retrieve. The mirroring behaviour can thus give a clue to the affix retrieval process.

Abbreviations

ABL	ablative	CAUS	causative
ACC	accusative	COM	comitative
ADJ	adjectivizer	CONTR	contrast
ADVZ	adverbializer	CVB	converb
ALIEN	alienable possession	DAT	dative
ALL	allative	DESIG	designative
AND	andative	DIS	disjunctive particle
ANT	anteriority	DUR	durative
CAR	caritive	EQT	equative

EXCL	exclusive	ONOM	onomatopoetic expression
FOC	focus particle	PASS	passive
FUT	future tense	PERF	perfective
GEN	genitive	PH1, PH2	placeholder stems
HAB	habitual	PL	plural
HES	hesitative	POSS	possessive
IMPER	imperative	PROL	prolative
INCH	inchoative	PROPR	proprietary
INCL	inclusive	PRS	present tense
INDEF	indefinite	PST	past tense
INS	instrumental	PTCL	particle
INTER	interrogative particle	PTCP	participle
INTJ	interjection	REP	repetitive
INTS	intensifier	RES	resultative
IPFV	imperfective	RFL	reflexive
LIM	limitative	SG	singular
LOC	locative	TR	transitive
NEG	negation	VBLZ	verbalizer
NFUT	non-future tense	R	Russian
NMLZ	nominalizer	SAH	Sakha
OBL	oblique stem		

References

- Aralova, Natalia & Brigitte Pakendorf. 2023. Non-canonical possessive constructions in Negidal and other Tungusic languages: A new analysis of the so-called “alienable possession” suffix. *Linguistics* 61(6). 1563–1592. DOI: 10.1515/ling-2022-0030.
- Bel’dy, Raisa A. & Tatyana D. Bulgakova. 2012. *Nanayskiye skazki [Nanai folk tales]*. Sibirien: Verlag der Kulturstiftung Sibirien. 268.
- Bulatova, Nadežda Ja. 2003. Mestoimeniya v sibirskoy gruppe tunguso-manchzhurskikh yazykov [Pronouns in the Siberian language group of the Tungusic family]. *Acta Linguistica Petropolitana* 1(3). 171–236.
- Bulatova, Nadežda Ja. & Lenore A. Grenoble. 1999. *Evenki* (Languages of the World/Materials 141). Munich: Lincom Europa.
- Cheung, Lawrence Yam-Leung. 2015. Uttering the unutterable with *wh*-placeholders. *Journal of East Asian Linguistics* 24. 271–308.
- Cincius, Vera I. 1982. *Negidal’skij jazyk [The Negidal language]*. Leningrad: Nauka.

- Cincius, Vera I. 1975–1977. *Sravnitelnyy slovar tunguso-manchzhurskikh yazykov: Materialy k etimologicheskому словарю [A comparative dictionary of Tungusic languages]*, vol. 2. Leningrad: Nauka.
- Däbritz, Chris Lasse & Valentin Yu. Gusev. 2021. *INEL Evenki Corpus. Version 1.0. Publication date 2021-12-31. Archived at Universität Hamburg. [Https://hdl.Handle.Net/11022/0000-0007-F43C-3](https://hdl.Handle.Net/11022/0000-0007-F43C-3). In: The INEL corpora of indigenous Northern Eurasian languages. [Https://hdl.Handle.Net/11022/0000-0007-F45A-1](https://hdl.Handle.Net/11022/0000-0007-F45A-1). DOI: 10.25592/uhhfdm.9628. <https://inel.corpora.uni-hamburg.de/EvenkiCorpus/search> (20 December, 2023).*
- Egorova, Anastasia D. 2017–2021. *Kur-Urmi Nanai language corpus*. <http://nanai.web-corpora.net> (20 December, 2023).
- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Gusev, Valentin Yu. & Svetlana Yu. Toldova. 2017–2021. *Field notes on the Uilta language*. <https://languagedepot.org/project/oaa-flex> (20 December, 2023).
- Gusev, Valentin Yu. & Daria O. Zhornik. 2023. *Uilta language corpus*. <http://uilta.web-corpora.net> (20 December, 2023).
- Hlavac, Jim. 2011. Hesitation and monitoring phenomena in bilingual speech: A consequence of code-switching or a strategy to facilitate its incorporation? *Journal of Pragmatics* 43(15). 3793–3806.
- Hölzl, Andreas. 2018. *A typology of questions in Northeast Asia and beyond: An ecological perspective*. Berlin: Language Science Press.
- Hölzl, Andreas & Thomas E. Payne (eds.). 2022. *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32). Berlin: Language Science Press. DOI: 10.5281/zenodo.7025328.
- Idiatov, Dmitry. 2007. *A typology of non-selective interrogative pronominals*. Antwerp: University of Antwerp. (Doctoral dissertation).
- Ikegami, Jiro. 2010. *Uilta oral literature : A collection of texts. Part 2* (Tsungūsu gengo bunka ronshū [Tungus Language and Culture Collection] 16). Hokkaido: Ōsakagakuindaiigaku jōhō gakubu [Osaka Gakuin University, Faculty of Informatics]. 107–137. <https://eprints.lib.hokudai.ac.jp/dspace/handle/2115/56959>.
- Kazakevich, Olga A., Elena Klyachko & Natalia Mitrofanova. 2023. *Siberian-Lang corpus*. <http://gisly.net/corpus> (20 December, 2023).
- Kazakevich, Olga A., Konstantin Polivanov, Anastasia Gruzdeva, Kuznetsov Nikita Aleksandrovich, Daria Mordashova, Evgenia Renkovskaya, Natalia Kar-dobolskaya, Zaire Khalilov, Sheipher Karina Olegovna & Sofia Ganieva. 2023.

- Minority languages of Russia. Ulta.* <https://minlang.iling-ran.ru/lang/uiltinskiy> (20 December, 2023).
- Kazama, Shinjiro. 1993. *Nanai-go tekisuto [Nanai texts]*. Otaru: Otarushōkadaigaku gengo sentā [Otaru University of Commerce Language Center]. <http://coe.aa.tufts.ac.jp/tungus/texts.html>.
- Kazama, Shinjiro. 1996. *Orochigo kiso shiryō [Oroch language basic materials]* (Tsungūsu gengo bunka ronshū [Tungus Language and Culture Collection] 7). Tottori: Tottoridaigaku kyōiku gakubu [Tottori University Faculty of Education]. http://coe.aa.tufts.ac.jp/tungus/orochi_kazama1996.
- Kazama, Shinjiro. 2007. Rhetorical questions with interrogative markers in Nanai. In Yuji Kawaguchi, Toshihiro Takagaki, Nobuo Tomimori & Yoichiro Tsuruga (eds.), *Corpus-based perspectives in linguistics*, 319–335. Amsterdam: John Benjamins. DOI: 10.1075/ubli.6.22kaz.
- Khasanova, Marina M. & Aleksandr M. Pevnov. 2003. *Myths and tales of the Negidals* (ELPR Publication Series A2-024). Osaka: Nakanishi.
- Kim, Minju. 2020. Korean general extenders *tunci ha* and *kena ha* ‘or something’: Approximation, hedging, and pejorative stance in cross-linguistic comparison. *Pragmatics* 30(4). 557–585.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hödl & Thomas Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- Kogura, N. 呂倉, 德和. 2021. Sibe text: Dancing of Sibe People. *Hoppō gengo kenkyū* 北方言語研究 [Tungusic languages research] 11. 213–232.
- Kogura, Norikazu. 2022. A narrative text from the Family problems picture task: Sibe (sjo). *Asian and African Languages and Linguistics* 16. 115–131.
- Malchukov, Andrej L. 1995. *Even* (*Languages of the world/materials* 12). München: München: LINCOM Europa. 44.
- Matić, Dejan. 2008. A note on hesitative forms in two even dialects. In O. A. Osipova, A. V. Didenko & Elena A. Krjukova (eds.), *Sravnitel'no-istoričeskoe i tipologičeskoe izučenie jazykov i kul'tur [Comparative-historical and typological studies of languages and cultures]: Materialy meždunarodnoj naučnoj konferencii XXV Dul'zonovskie čtenija* [The materials of the 25th international scientific conference Dulzon Readings], 234–239. Tomsk: Tomsk State University.
- Mauri, Caterina & Andrea Sansò. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. *Folia Linguistica* 52. 1–35. DOI: 10.1515/flih-2018-0001.
- Nakagawa, Hiroshi. 2023. *Sound materials of the Nivkh language*. <http://ext-web.edu.sgu.ac.jp/hidetos/HTML/SMNStitle.html> (20 December, 2023).

- Nikolaeva, Irina, Edward Garrett & Oliver Bond. 2019. *Endangered languages and cultures of Siberia*. <http://www.siberianlanguages.surrey.ac.uk> (20 December, 2023).
- Nikolaeva, Irina & Maria Tolskaya. 2001. *A grammar of Udihe*. Berlin: De Gruyter Mouton.
- Oskolskaya, Sofia, Ezequiel Koile & Martine Robbeets. 2022. A Bayesian approach to the classification of Tungusic languages. *Diachronica* 39(1). 128–158.
- Overstreet, Maryann. 1999. *Whales, candlelight, and stuff like that: General extenders in English discourse*. Oxford: Oxford University Press on Demand.
- Overstreet, Maryann. 2020. The English general extender: The forms and functions of a new linguistic category, or something, and stuff. *English Today* 36(4). 47–52.
- Overstreet, Maryann & George Yule. 2021. *General extenders: The forms and functions of a new linguistic category*. Cambridge: Cambridge University Press.
- Ozoliņa, Larisa V. & Boris V. Boldyrev. 2001. *Oroksko-russkiy slovar [Ulta-Russian dictionary]*. Novosibirsk: SO RAN.
- Pakendorf, Brigitte. 2025. A corpus-based investigation of fillers in Negidal (Northern Tungusic). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitations and placeholders*, 75–129. Berlin: Language Science Press. DOI: 10.5281/zenodo.1569759.
- Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct Northern Tungusic language of the Lower Amur* (<http://hdl.handle.net/2196/00-0000-0000-000F-E387-1>). London: SOAS, Endangered Languages Archive.
- Pakendorf, Brigitte, Dejan Matić, Natalia Aralova & Alexandra Lavrillier. 2010. *Documentation of the dialectal and cultural diversity among Evens in Siberia. Nijmegen & Leipzig: Dokumentation Bedrohter Sprachen/Max Planck Institute for Psycholinguistics/Max Planck Institute for Evolutionary Anthropology*. <https://Hdl.Handle.Net/1839/00-0000-0000-000F-2DFE-0> (20 December, 2023).
- Panfilov, Vladimir Z. 1965. *Grammatika nivkhskogo yazyka. Chast I [The grammar of Nivkh. Part I]*. Moscow: AN USSR.
- Petrova, Taisija I. 1936. *Ulchskiy dialekt nanayskogo yazyka [The Ulch dialect of Nanai language]*. Moscow: Uchpedgiz.
- Petrova, Taisija I. 1967. *Yazyk orokov (ulta) [The Ulta language]*. Leningrad: Nauka.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.

- Robbek, Vasilij A. & Maria E. Robbek. 2005. *Evensko-russkij slovar'* [Even-Russian dictionary]. Novosibirsk: Nauka.
- Rose, Françoise. 2025. One more thing ‘thing’ can do in Tupí-Guaraní languages: The Teko filler. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 441–484. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697597.
- Sakhatyla. 2023. *Sakhatyla. Sakha online dictionaries*. <https://sakhatyla.ru> (20 December, 2023).
- Savelyeva, Valentina N. & Chuner M. Taksami. 1970. *Nivkhsko-russkiy slovar': Svyshe 13 000 slov* [Nivkh-Russian dictionary: More than 13 000 words]. Moscow: Sovetskaia: Sovetskaya entsiklopediya.
- Seraku, Tohru. 2024. Placeholders in crosslinguistic perspective: Abilities, preferences, and usage motives. *Linguistics* 62(2). 1–26. DOI: 10.1515/ling-2023-0068.
- Shakhovtsov, Kirill G., Dmitry Anatolyevich Funk, Mamontova Nadezhda Alexandrovna & Alexandra Nikolaevna Terekhin. 2023. *Corpora at IEA RAS*. <http://corpora.iea.ras.ru/corpora/structure.php> (20 December, 2023).
- Sharina, Sardana I. & Raisa P. Kuzmina. 2018. *Nizhnekolymskii govor evenskogo jazyka*. [Lower Kolyma dialect of the Even language]. Novosibirsk: Izd-vo SO RAN.
- Stoynova, Natalia M. 2017–2021. *Ulch language corpus*. <http://ulch.web-corpora.net> (20 December, 2023).
- Šneider, Evgenij R. 1936. *Kratkij udejsko-russkij slovar' s priloženijem gramatičeskogo očerka* [Concise Udihe-Russian dictionary with a sketch on grammar]. Moscow: Učpedgiz.
- Šneider, Evgenij R. 1937. *Materialy po jazyku anjujskix ude* [Materials on the language of the Aniuj Udihe]. Moscow: Naučno-issledovatel'skaja asociacija Instituta Narodov Severa im. I.G. Smidoviča.
- Tárnyiková, Jarmila. 2019. English placeholders as manifestations of vague language: Their role in social interaction. *Brno studies in English* 45(2). 201–216.
- Tolskaya, Maria & Inna Tolskaya. 2008. Question particles or what? Open alternative questions in Udeghe. In Edward J. Vajda (ed.), *Subordination and coordination strategies in North Asian languages*, vol. 300, 97. Amsterdam: John Benjamins Publishing.
- Vasilevich, Glafira M. 1958. *Evenkijsko-russkiy slovar'* [Evenki-Russian dictionary]. Leningrad: Gosudarstvennoye izdatelstvo inostrannykh i natsionalnykh slovarey.
- Ventayol-Boada, Albert. 2025. Copula or placeholder: A qualitative and quantitative analysis of *λe-* in Kolyma Yukaghir discourse. In Brigitte Pakendorf &

- Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 131–168. Berlin: Language Science Press. DOI: ??.
- Visser, Eline. 2025. Form and frequency of Kalamang fillers. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 251–273. Berlin: Language Science Press. DOI: ??.
- Yamada, Yoshiko. 2011. Uiruta no fu gakki ‘yōdopu’ : Fukugen kōtei ni tsuite no uiruta-go kita hōgen tekisuto o chūshin ni [A type of shaken musical instruments joodopu of the Uilta: With a text in Uilta Northern dialect representing process of restoration]. *Hoppō jinbun kenkyū [Northern humanities research]* 4. 25–49.
- Yamada, Yoshiko. 2015. Uiruta-go kita hōgen tekisuto: Hito kui obake no hanashi [An Uilta text in Northern dialect: A story about an ogre]. *Hoppō gengo kenkyū [Tungusic languages research]* 5. 261–280.
- Yamada, Yoshiko. 2017. Uiruta-go kita hōgen no on'in-teki keitai-teki tokuchō: Minami hōgen to no sōi-ten o chūshin ni [Some phonological/morphological features from the Northern dialect of Uilta: Especially on differences from the Southern dialect]. *Hoppō jinbun kenkyū [Northern humanities research]* 10. 51–70.
- Zikmundová, Veronika. 2013. *Spoken Sibe: Morphology of the inflected parts of speech*. Prague: Karolinum Press.

Chapter 3

A corpus-based investigation of fillers in Negidal (Northern Tungusic)

Brigitte Pakendorf

Laboratoire Dynamique du Langage (CNRS & Université Lumière Lyon 2)

The Northern Tungusic language Negidal, which is spoken in the Russian Far East by barely a handful of elderly speakers, is a language with a dedicated filler, namely the root *ujun*. As shown by an investigation of a corpus of annotated oral recordings comprising over 200 texts of diverse genres (Pakendorf & Aralova 2017), this filler is frequently used: over 880 tokens of *ujun* are found in ~76,000 Negidal words, amounting to nearly 12 tokens/1000 words; however, there are large inter-speaker differences. About 70% of the occurrences of the filler are used as placeholders for either nouns or verbs, and occasionally even clauses or direct speech. Targets are frequently omitted; when present, it is inflectional, but not derivational, morphology that is mirrored on the placeholder. This mirroring demonstrates that speakers plan the syntactic frame of their utterance well in advance, with differences between less and more proficient speakers. Although the filler is used when speakers have difficulties accessing a word, that is not the only factor triggering its use, as shown by the fact that as a placeholder it often stands in for common lexemes or items that were mentioned in the immediately preceding discourse. The Negidal data add to our knowledge of the as-yet understudied domain of fillers and open up further questions concerning the management of disfluencies in Negidal.

keywords: disfluency, hesitation, placeholder, speech planning

1 Introduction

Speakers who experience problems with the “planning and execution of speech” (Lickley 2015: 452) have various means at their disposal to signal their difficulties and to play for time while searching for an elusive word or planning how to



continue their discourse, such as pausing or drawing out final syllables. Another option is to use a **FILLER**, an overt marker of hesitation with a conventionalized form (Clark & Fox Tree 2002). Two common strategies – not necessarily mutually exclusive – for this type of delay function can be distinguished: **PLACEHOLDERS** that substitute for a particular word (the target item) and that are thus fully integrated referential expressions, and **HESITATIVES**, non-referential markers of hesitation that are not syntactically integrated (Hayashi & Yoon 2006). Such fillers often develop out of demonstratives (Hayashi & Yoon 2006), indefinite pronouns, interrogative pronouns, or semantically bleached nouns, but can also lexicalize out of phrases, such as English *whatchamacallit* (Podlesskaya 2010: 12–13).

Negidal (ISO-639: neg, glottocode: negi1245), a Northern Tungusic language of the Lower Amur region, has a dedicated filler *uŋun* (glossed **FLR** in all the examples) with opaque etymology (however, see Klyachko 2025 [this volume] for a possible origin as an interrogative pronoun). This can function as both a placeholder (1, 2) and a hesitative (3). In its placeholder use *uŋun* can stand in for both nouns (1) and verbs (2).

- (1) *gə əmə-tca-βun* (2050) ***uŋu-l-βəj*** (2770) ***tigə-l-βəj***
 DP come-PST-1PL.EX FLR-PL-PRFL.PL dish-PL-PRFL.PL
na:-tca-βun taj ugi-la
 put-PST-1PL.EX DIST top-LOC
 ‘Well, we arrived and put **our whatchamacallits**, our dishes at the top.’
 [GIK_2tatarskoe: 112; ID 732]
- (2) *taj komalan-ma tčak-mi* ***ti:*** (0) ***uŋun-ə-s*** (0)
 DIST fur.rug-ACC gather-ss.COND like.this FLR-NFUT-2SG
atal-i-s
 take.off-NFUT-2SG
 ‘That fur rug... gathering **you whatchamacallit**, you take off (the scalp)
 like this.’ [DIN_duck_heads: 14; ID 571]
- (3) *i:-ji-ji-j...* ***i:-ji-ji-j*** ***gəsə***
 enter-PRS.PTCP-INS-PRFL.SG enter-PRS.PTCP-INS-PRFL.SG together.NNC
 (2180) ***uŋuna*** (1760) *su:n-mi bu:-jə-n loko-da-tin*
 FLR coat-PRFL.SG give-NFUT-3SG hang.up-VS.PURP-3PL
 ‘As soon as she entered... **uhm**... she gave her coat for (them) to hang it.’
 [DIN_restoran: 4; ID 671]

The examples are sourced with the text identifier plus the number of the annotation unit and can be found in Pakendorf & Aralova (2017). In addition, the

unique ID from the coding sheet that underlies the analysis presented in this paper is provided; the coding sheet and further supplementary material can be found in the Negidal fillers dataset: <https://hdl.handle.net/11403/negidal-filters>. Note that I add ‘whatchamacallit’ – for both nominal and verbal uses – or ‘uhm’ to the translations of examples in order to facilitate their understanding, even when the filler was not translated by our consultants. Furthermore, I add the length of pauses preceding and following the filler to the text line in parentheses, in milliseconds rounded to tens and with ‘no pause’ being shown as 0. In examples with more than one filler, it is only the relevant filler that is marked in this way; this is furthermore highlighted in bold, with the target, where relevant, underlined.

In their discussion of demonstratives as fillers, Hayashi & Yoon (2006: 488) distinguish three “related but distinct usage types”, namely placeholders, avoidance markers, and hesitations. I did not identify any obvious avoidance use¹ for the Negidal filler *uŋun*, and given the highly culture-specific and context-sensitive nature of such items, and without the possibility to discuss the data with speakers, I feel I would not be reliably able to do so. This article therefore focuses on placeholder and hesitant uses of the filler *uŋun* in Negidal in order to enrich our knowledge of this as yet understudied domain. It has two major goals: to describe the morphosyntactic characteristics of the filler and to investigate the cognitive and interactional aspects of its use.

My study is based on a corpus of transcribed, translated, and glossed oral recordings (Pakendorf & Aralova 2017; see §2 for details), with the preliminary typology established by Podlesskaya (2010) as my theoretical point of departure. This was considerably extended, however, by the novel insights found in the other chapters included in this volume. The rest of the chapter is structured as follows: the data and methodology of the study are introduced in §2, and the distinction (or lack thereof) between the placeholder and hesitant strategy of the filler is addressed in §3. The subsequent section is dedicated to a detailed description of the Negidal filler, with various morphosyntactic characteristics outlined in §4.1 and the differences in use between speakers discussed in §4.2. The pragmatic and cognitive aspects of use of the filler are the topic of §5, and the paper concludes with a brief discussion in §6.

¹This refers to the use of fillers to replace words that are socially or culturally sensitive, use of which might be perceived as impolite or offensive (Hayashi & Yoon 2006: 501–507).

2 Data and methodology

Negidal, a close sister of the better-known and better-described language Evenki (Oskolskaya 2024; own knowledge), is spoken on the Lower Amur river and its tributary, the Amgun', in the Russian Far East. Although two dialects, Upper and Lower Negidal, are known (Myl'nikova & Cincius 1931, Pevnov & Khasanova 2006: 453), nowadays at most a handful of elderly women speak the Upper dialect, and the Lower dialect is extinct (Kalinina 2008: 272, Pakendorf & Aralova 2018). It is a morphologically rich language (Pakendorf & Aralova 2020), with nominal inflectional morphology comprising nine cases (including the unmarked Nominitive) and possessive suffixes that index the person and number of the possessor;² singular number is unmarked as opposed to the marked plural. Verbs carry diverse tense, aspect, mood, and valency-changing suffixes as well as obligatory subject indexes; they are negated with analytical constructions consisting of the inflected negative auxiliary *ə-* and the lexical verb carrying an invariant suffix (here glossed as NEG.CVB ‘Negative Converb’; cf. Hözl 2015). Common nominal derivational suffixes include the Diminutives *-kan* and *-tcan* and the “Decessive” suffix (which designates deceased individuals) *-ŋasa*; frequent verbal derivational suffixes include the Associated Motion suffix *-na* and various Aktionsart markers, such as the Accelerative *-maltca* or the Multiplicative *-kta*. Although objects are not indexed, both subjects and objects are frequently omitted in spontaneous speech.

The recordings on which this study is based were made in the village of Vladimirovka (Polina Osipenko district, Khabarovsk Krai) at different times by different individuals (Table 1): in the 1990s, a female speaker (TIN) taped herself reading or reciting excerpts of narratives for use as little lessons of Negidal broadcast by the regional radio station, and a male speaker (APK) telling folktales was taped by a linguist and, on a different occasion, by an ethnologist; these different cassette tapes³ were subsequently digitized for a large documentation project (Pakendorf & Aralova 2017). Between 2005 and 2010 the oldest speaker still alive at the time (APN) and three of her daughters (DIN, GIK, LIO) were recorded by a team of Russian linguists (Kalinina 2013), and in 2017 and 2020 further recordings were made by Natalia Aralova and myself (Pakendorf & Aralova 2017). The codes used to refer to the speakers are abbreviations of the speakers’ first name, patronymic, and surname.

²These are glossed only with the person and number, not with a separate gloss to indicate the possession (i.e. *ge:-β* ‘my friend’ is glossed as friend-1SG and not friend-px.1SG).

³Unfortunately, we have absolutely no metadata for any of these tape recordings.

Table 1: Some characteristics of the text corpus

Speaker	Sex	Birthyear	Comment	# words
APN	F	1916		21,900
APK	M	1921		4,200
TIN	F	1940	daughter of APN	7,800
DIN	F	1942	daughter of APN	29,000
GIK	F	1945	daughter of APN	8,500
AET	F	1947		2,500
LIO	F	1948	daughter of APN	600
AVK	F	1950	niece of APK	960
GIKlju	F	1955		310

Overall, over 18 hours of recordings have been transcribed, translated, and glossed; these comprise over 200 texts of diverse genres (mainly folklore and real-life anecdotes, but also conversations and procedural texts) amounting to ~78,000 words (~76,000 excluding larger switches to Russian) and 14,100 annotation units. The corpus is far from being balanced, since only nine speakers are represented, eight women and only one man (APK); of these, only four are still alive at the time of writing. Furthermore, the contributions of individual speakers are far from equal, and five of the represented speakers are closely related, namely APN and four of her daughters (Table 1). All of the speakers used or use languages other than Negidal regularly in daily life: Evenki and Russian for APK and APN (who in addition spoke Nanai during her childhood and adolescence), and Russian for the younger speakers. The source of lexemes or morphemes of non-Negidal origin is indicated in the glosses with the following abbreviations: Evk: Evenki; Nnc: Nanaic (i.e. a lexeme that could have its source in Nanai, Ulch, or Orok/Uilta); R: Russian; Y: Sakha (also known as Yakut).

The filler *uŋun* was extracted via concordance search in the FieldWorks⁴ database in which the texts have been glossed. In order to verify the reliability of the transcription as well as to obtain prosodic information, each utterance was checked by listening to the recording in ELAN⁵ and subsequently coded for diverse information on the morphosyntax of the filler and its target, where relevant, as well as on various aspects of its use (see the coding definitions deposited with

⁴FieldWorks Language Explorer version 9.1, SIL International.

⁵ELAN Linguistic Annotator version 6.5, MPI for Psycholinguistics, Nijmegen. Retrieved from <https://archive.mpi.nl/tla/elan>

the Negidal fillers dataset⁶ for details). In total, 882 tokens of *ujun* were coded in this way and then analysed in Excel.

Since assigning a strategy to the filler or identifying which element might be the target of a particular placeholder was very difficult for many of the tokens, in a second step I investigated the intonation pattern of all the utterances containing fillers using Praat.⁷ For this, my colleague Jennifer Krzonowski wrote a script that cut each utterance out of its recording context and annotated it with the transcription in one tier. I myself then added a second tier with the location of the filler and the position and duration of surrounding pauses, where relevant. In a final step, the script automatically drew the waveform plus pitch contour.⁸ The intonation patterns were inspected individually to identify the strategy of each token of the filler, with the larger context, the translation provided by our consultants in the field and notes from clarification sessions on what form the filler should have taken, as well as clues from video recordings taken into account where possible. Nevertheless, it was impossible to assign a strategy to 24 tokens (2.7% of the total), either because the context was too unclear or because they carried morphology that did not make any morphosyntactic sense; these were coded as ‘unclear’. Furthermore, as will be discussed in the following section, it is impossible to always distinguish between the hesitative and the placeholder strategy, and 50 tokens (5.7%) were coded as ‘indeterminate’, following the terminology of Billings & McDonnell 2025 [this volume].

3 Placeholders vs. hesitations

As mentioned in the introduction, the distinction between placeholders and hesitations is assumed to lie in the syntactic integration of the filler: hesitations are not morphologically integrated, can occur anywhere in the sentence, and do not have referential use (Hayashi & Yoon 2006: 512), while placeholders are morphologically and syntactically fully integrated referential expressions (Hayashi & Yoon 2006: 490). However, Hayashi & Yoon (2006) base their analysis primarily on Japanese and Korean, where distinct elements express the two strategies,

⁶<https://hdl.handle.net/11403/negidal-filters>

⁷Praat version 6.1.37, Paul Boersma and David Weenink.

⁸Note that the figures in this chapter were produced manually by myself in order to be able to adjust the annotation and font size; in addition, whereas the Praat script worked with the sound files extracted from the videos, if these exist, I based the figures on the original sound file as recorded with a Marantz or Zoom audio recorder to maximize the quality. The data deposited in the Negidal fillers dataset (<https://hdl.handle.net/11403/negidal-filters>) hence occasionally contain two sound files and Textgrids for a token of the filler.

with pronominal demonstratives functioning as placeholders, while adnominal demonstratives function as hesitatives. Since in Negidal a single lexeme fills both strategies, it is not always easy to identify which strategy is manifested by a given occurrence of the filler – a problem also found for French (Hennecke & Mihatsch 2022), Northern Pastaza Kichwa (Rice 2025 [this volume]), and the Austronesian languages Besemah and Nasal (McDonnell & Billings 2025 [this volume]; Billings & McDonnell 2025 [this volume]).

Some tokens of the filler can be easily identified as hesitatives, such as (3), repeated here for convenience as (4): the pitch and the waveform intensity are low (Figure 1), the filler is surrounded by long pauses, and hesitation is furthermore manifested in the repetition of the participle at the beginning of the sentence.

- (4) *i-ji-ji-j... i-ji-ji-j gəsə*
 enter-PRS.PTCP-INS-PRFL.SG enter-PRS.PTCP-INS-PRFL.SG together.NNC
 (2180) *ujunə* (1760) *su:n-mi bu:-jə-n loko-da-tin*
 FLR coat-PRFL.SG give-NFUT-3SG hang.up-VS.PURP-3PL
 ‘As soon as she entered... uhm... she gave her coat so that they would
 hang it up.’ [DIN_restoran: 4; ID 671]

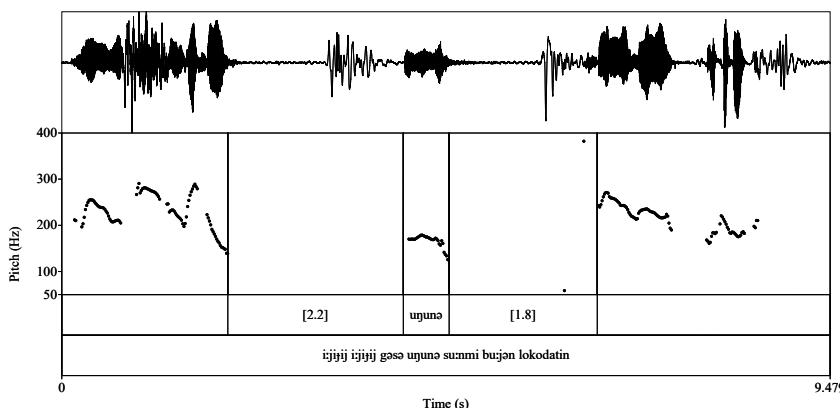
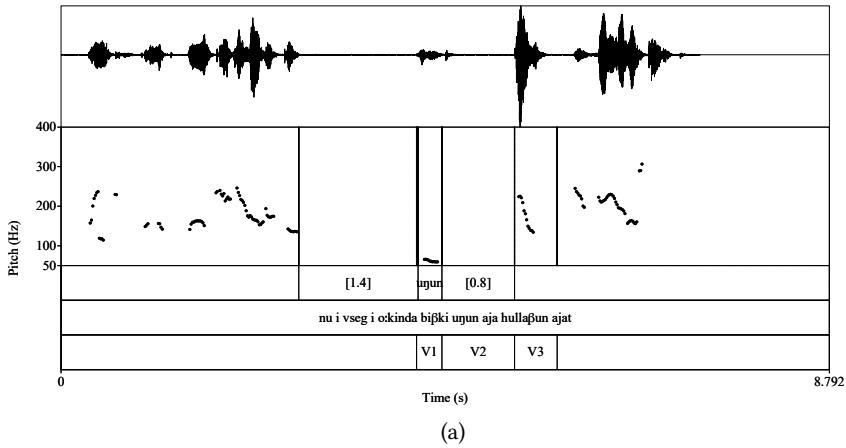


Figure 1: Waveform and pitch contour of example (4)

Similarly, there can be no doubt that the filler in (5) is a hesitative: it is surrounded by fairly long pauses, the pitch drops down, and the intensity of the waveform is very low (Figure 2a). In addition, in the accompanying video (Figure 2b) one can see that the speaker, who is wearing a striped shirt and sits facing the camera, looks to the left after pronouncing *ujun* and then turns back to face the camera as she produces the word that had eluded her, *aja(t)* ‘good, well’. Such

gaze aversion that accompanies hesitation uses of the filler has been described for Swedish (Allwood et al. 2005), Dutch (Jehoul et al. 2016), and Northern Pastaza Kichwa (Rice 2025 [this volume]). It has been suggested that it allows the speaker to “concentrate on the organization of the utterance” while also being a signal of “his intention to continue to hold the floor” (Kendon 1967: 42) and is characteristic of conversations rather than of monologic situations such as class presentations (Kosmala 2022).

- (5) *nu i vseg... i o:kin=da bi-βki (1420) uyun (900)*
 PTL.R and.R FS and.R when=ADD COP-HAB.PTCP FLR
aja hul-la-βun aja-t
 good go-NFUT-1PL.EX good-ADVR
 ‘And we always... uhm... go well.’ [AVK_crossing_stream: 142; ID 502]



(a)



(b)

Figure 2: (a) Speech waveform and pitch contour of example (5). The annotation “V1, V2, V3” indicates the approximate timepoint of the video stills shown below the intonation pattern. (b) Three still frames taken from the video recording of example (5).

In contrast, instances in which the filler carries morphology can often be straightforwardly recognized as having a placeholder function, such as in (2),

repeated here as (6): the filler carries the same Non-Future tense marker and 2SG subject index as the target *atalis* ‘you take off’, there are no pauses surrounding the filler, and the pitch and the waveform intensity of *uŋun* are as high as those of the rest of the utterance (Figure 3; note that the pitch line of the first syllable of *atalis* is erroneously missing).

- (6) *taj komalan-ma tca-k-mi ti: (0) uŋun-ə-s (0)*
 DIST fur.rug-ACC gather-ss.COND like.this FLR-NFUT-2SG
atal-i-s
 take.off-NFUT-2SG
 ‘That fur rug... gathering you whatchamacallit, you take off (the scalp)
 like this.’ [DIN_duck_heads: 14; ID 571]

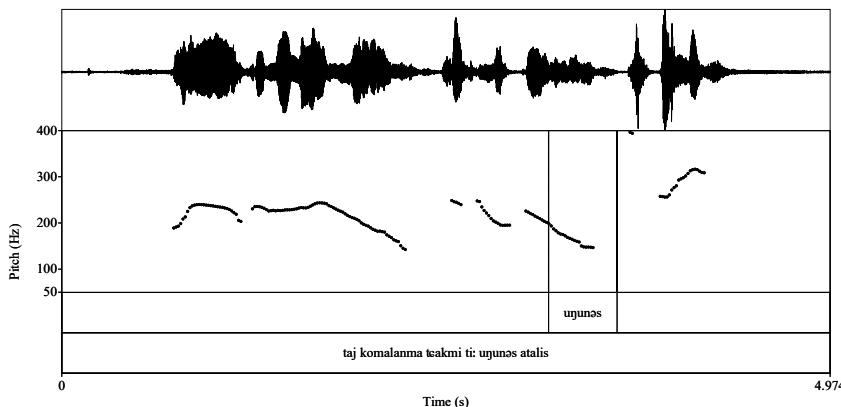
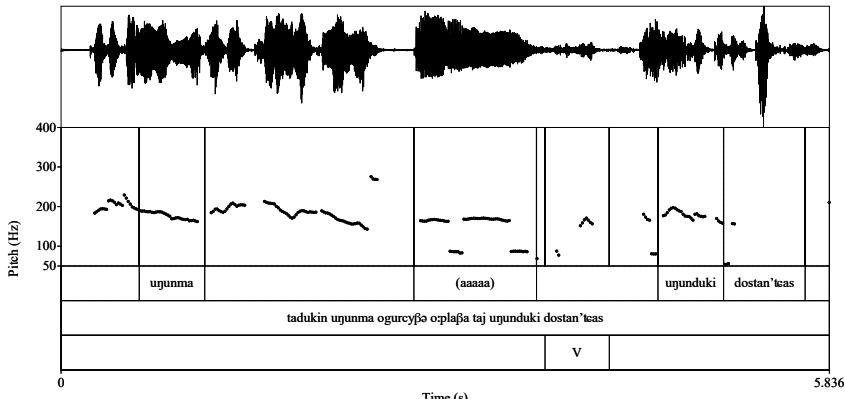


Figure 3: Waveform and pitch contour of example (6)

The two instances of *uŋun* in (7) are equally easy to identify as placeholders: they carry nominal case suffixes, there is no pause, and the pitch and waveform intensity are at the same level as those of the rest of the utterance (Figure 4a). Note that the element *aaaa* entered in brackets is an interjection by the speaker’s interlocutor, and that the verb *dostan’tcas* ‘you got hold of something’ is not reflected in the pitch contour. While the first placeholder is followed by its target *ogurcyβə* ‘cucumbers’, the second one holds the place of the intended target ‘cellar’ by itself. It is preceded by a pointing gesture in the general direction of the cellar opening (Figure 4b). A similar pointing gesture accompanies the Komnzo placeholder in its use to fill in for a place name (Döhler 2025 [this volume]).

- (7) *tadukin* (0) *ujun-ma* (60) *ogurcyβə* *o:-pla-βə* *taj*
 then FLR-ACC cucumbers.R-ACC make-PASS.PTCP-ACC DIST
 (0) *ujun-duki* (0) *dostan'-tca-s*
 FLR-ABL get.R-PST-2SG
- 'Then you got whatchamacallit, marinated cucumbers from that
 whatchamacallit [the cellar].'
 [AVK_GIK_wine: 12; ID 525-526]



(a)



(b)

Figure 4: (a) Waveform and pitch contour of example (7). The annotation "V" shows the approximate timepoint of the still frame shown below the intonation pattern. (b) Still frame taken from the video recording of example (7).

However, matters are far from always being so straightforward. While morphologically marked fillers tend to function as placeholders, this is not in itself a

sufficient criterion to identify an instance of filler use as a placeholder, because the hesitatively used filler sometimes ends in an epenthetic schwa or [e] (e.g. (4)), making it identical to both the Non-Future 3PL and the Negative Converb form of the verb, *ujunə*. This is illustrated by the following, very long example (8), where there are four fillers, all with identical form, *ujune*, and no clear distinctions in pitch or intensity of the waveform (Figure 5), albeit with longer pauses after the first and third token than after the second and fourth token. Note that due to the length of the example I exceptionally show only the location of the filler (abbreviated as 'flr' in the figure) on one tier and the location and duration of the surrounding pauses on the second tier. The second and fourth token are identifiable as verbal placeholders, as shown by the fact that they follow upon the negative auxiliary *odan* 'so that it (does) not' and precede their targets, the negated verbs *alilla* '(not) get angry' and *anuttce* '(not) push away'. The first and third token, in contrast, are likely to be hesitatives, since they are not syntactically integrated. The first one comes at a point where the speaker decides to restructure her sentence: after starting off with what seems to have been intended as a definition of the concept *holgin*, she switches to the 1PL.EX subject pronoun *bu* 'we', but then starts over with *oj bisi boga* 'this place where (we) live' as the subject. The third token of *ujune* intervenes in the clause *o:n gunə sagdil, boga nojanman gujandan* 'as the old people say, in order for the earth to take pity (on people)'. This shows clearly that even though they are similar in both intonation and form, these tokens of the filler represent two different strategies of filler use.

- (8) *taj holgi-ja holgin èto bu* (280) ***ujune*** (1690)
 DIST do.holgin-NFUT[3PL] holgin.ritual this.R 1PL.EX FLR
oj bi-si boga o-da-n (210) ***ujun-e*** (1000) *taj*
 PROX live-PRS.PTCP place NEG-VS.PURP-3SG FLR-NEG.CVB DIST
bəjə-tki ali-l-lə *ili o:n gunə sagdi-l* (0)
 person-ALL be.angry-INCH-NEG.CVB or.R how say-NFUT[3PL] old-PL
ujune (2760) *boga nojanman gujan-da-n o-da-n* (0)
 FLR nature 3SG.ACC pity-VS.PURP-3SG NEG-VS.PURP-3SG
ujun-e (560) *anut-tce*
 FLR-NEG.CVB push.away-NEG.CVB

'The holgin ritual is... we... uhm... so that our nature that we live in, so that it does **not whatchamacallit**, not get angry at people, as the old people say... uhm... so that the nature takes pity on people, so that it does **not whatchamacallit**, not push him away [from itself].'

[GIK_ostroe: 25; ID 778-781]

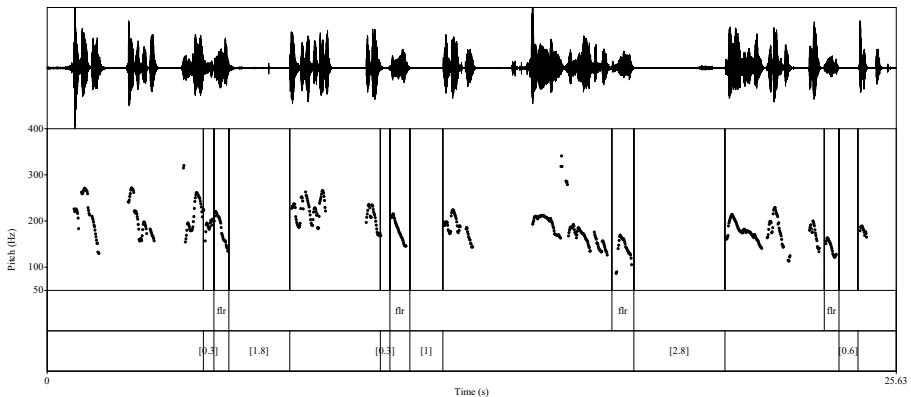


Figure 5: Waveform and pitch contour of example (8)

Furthermore, while the filler used as a hesitative stands in its bare form, the Nominative case in Negidal, which is the case taken by subjects, predicate nominals, or possessors, is also zero-marked. Hence morphologically unmarked tokens of the filler cannot automatically be interpreted as hesitatives, since a bare filler can be a placeholder that is substituting for a subject, predicate nominal, or possessor ((9) and Figure 6; cf. Rice 2025 [this volume], for a similar problem with morphologically unmarked tokens of the filler *mashti* in Northern Pastaza Kichwa). It can also stand in for noun phrases or adjectives in predicative position, and even for clauses or direct speech (see §4.1.4 and §4.1.5 for details).

- (9) *ne Pavel hutə-βa-n* (0) ***ujun*** (0) *hutə-βa-n*
 not.R PN offspring-ACC-3SG FLR offspring-ACC-3SG
 {APN:... the older son of Avujin pushed...} {DIN: whom, Pavel's child?}
 APN: 'not Pavel's child, whatchamacallit's child'

[APN_tri_sosed: 100; ID 467]

There are also a few instances when the bare filler appears to be used as a placeholder for a morphologically marked target (such as the Accusative case-marked noun 'soil' in (10)), a cross-linguistically not uncommon phenomenon (Podlesskaya 2010: 18–19). That *ujun* is used as a placeholder, and not a hesitative, in these instances is suggested by the fact that the pitch on the filler is as high as that of the preceding word, the intensity of the waveform is as strong as in the rest of the utterance, and there are practically no pauses (Figure 7).

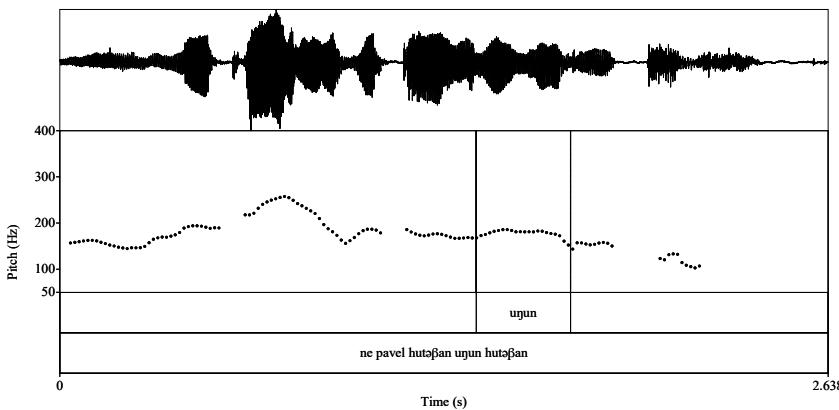


Figure 6: Waveform and pitch contour of example (9)

- (10) *o-ŋa:ti-s taj (0) **ugun** (80) tokala-βa tuŋku-ja
NEG-DEONT-2SG DIST FLR soil-ACC to.touch-NEG.CVB
ŋa:la-ji-j=da
hand-INS-PRFL.SG=ADD*
'You should not touch whatchamacallit, the earth with your hands...'
[GIK_zaprety: 30; ID 826]

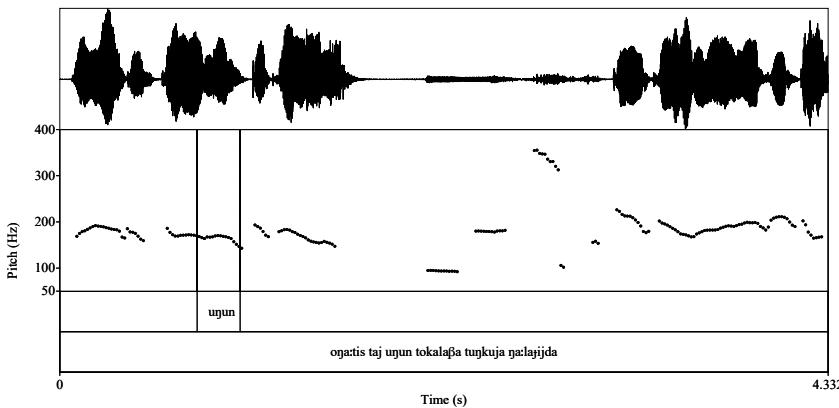


Figure 7: Waveform and pitch contour of example (10)

Like the form, intonation is also not always a sure indicator of the strategy of the filler, as shown by (11): here the filler carries the Accusative case found on the target and is thus identifiable as a placeholder, yet it is surrounded by

long pauses (the pause following the filler is so long that the clause containing the descriptive phrase substituting for the target was transcribed as a separate utterance, as indicated by the double slashes in the example). In addition, the pitch and the waveform intensity are very low (Figure 8). Without the nominal morphology this token would probably have been identified as a hesitative rather than as a placeholder.

- (11) *tajigdə petux uyun-du-j mi:je-du-j*
 DIST=CONTR cock.R FLR-DAT.ESS-PRFL.SG shoulder-DAT.ESS-PRFL.SG
hena-ja-n (700) uyun-ma (2990) //
 carry.on.back-NFUT-3SG FLR-ACC
mi:je-du-j hena-ja-n tco:ka-βa
 shoulder-DAT.ESS-PRFL.SG carry.on.back-NFUT-3SG grass-ACC
hosí-ββi sələ-βa
 cut(hair)-IMPERS iron-ACC
 ‘And that cock is carrying whatchamacallit on his whatchamacallit, on
 his shoulder. // On his shoulder he is carrying a scythe (lit. an iron for
cutting grass).’ [TIN_hare_fox: 89-90; ID 856]

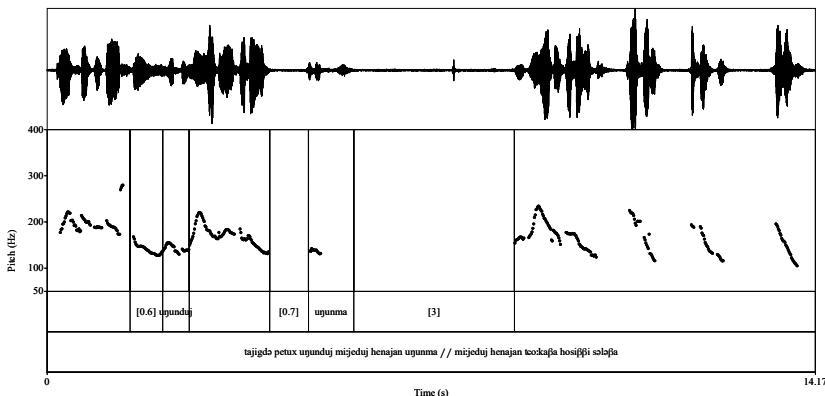


Figure 8: Waveform and pitch contour of example (11)

It was therefore not always possible to unambiguously assign bare tokens of the filler or those that end in a schwa to either the hesitative or the placeholder strategy.⁹ Such cases (50 in all) were coded as ‘indeterminate’.

⁹There is also one token of the filler (ID 486) that appears to be carrying the plural morpheme *-l*, which would partially mirror the morphology on a noun that might be the target. However, since the prosody was quite hesitative, and the speaker is APN whose enunciation is often rather indistinct, this was also coded as ‘indeterminate’.

These data indicate that in Negidal there is no clear-cut distinction between the hesitative and placeholder strategy of the filler, but that it is rather a single form with a single function of delaying speech production until a word or phrase that is eluding the speaker has been found. This accounts for the frequent ambiguous prosodic patterns, where clear placeholders, i.e. *uŋun* carrying morphology and followed by an overt target, are pronounced with low pitch and intensity and are accompanied with markers of hesitation such as silent or filled pauses or repetitions, while clear hesitatives that are not at all syntactically integrated can be pronounced with utterance-level intensity and pitch and without any pauses or other marks of hesitation. A similar lack of distinction between the hesitative and placeholder strategies of fillers have been found for French *truc* and *machin* (Hennecke & Mihatsch 2022), for the Nasal fillers (Billings & McDonnell 2025 [this volume]), and the fillers found in Besemah, where it is “often impossible to decide whether a filler is employed as a hesitator or placeholder” (McDonnell & Billings 2025 [this volume]).

4 The filler *uŋun* in use

4.1 The characteristics of the filler

4.1.1 Frequency, form, and function

Overall in the Negidal corpus there are 882 examples of *uŋun* in ~76,000 Negidal words,¹⁰ i.e. ~11.6 occurrences of the filler per thousand words. This is of comparable frequency as that found for the Evenki filler *anji* (~12.6/1000 words, Klyachko 2022: 213) and the Tagalog interrogative *ano* ‘what’ with “placeholder” and “filler” functions (~9.4/1000 words, calculated from data found in Nagaya 2022: 94, 105), but considerably more frequent than the Mandarin demonstrative-derived fillers *nage* and *zhege* (~6.7/1000 words; Zhao & Jurafsky 2005) or various Russian place-holders (~4/1000 words; Podlesskaya & Kibrik 2009: 186).

About two-thirds of the occurrences of the filler carry morphology (Table 2), with roughly equal numbers of overt nominal morphology and verbal morphology. Thirty-two examples of the bare filler are syntactically in the Nominative case (labelled NOM in the table), which, as mentioned in §3, is unmarked; they

¹⁰Since the filler occurs practically exclusively in Negidal contexts and not in Russian code-switching (with two exceptions), I base my calculation on the (approximate) number of Negidal words.

were kept separate from the unmarked tokens due to their distinct syntactic function. There is an overwhelming preponderance of inflectional morphology found on the filler, in contrast to derivational suffixes,¹¹ which make up only about 3% of the morphologically marked occurrences of the filler.

Table 2: Types of morphology found on the filler (in per cent). Unclear tokens are excluded.

unmarked	30.4
NOM	3.6
nominal	32.4
verbal	33.5

The filler mostly functions as a placeholder, with 71% of the 858 tokens for which a strategy was identified being assigned to this category, and only 23% being definable as hesitatives; approximately 6% are indeterminate. This differs from the filler use of Amazonian Spanish *este*, where the hesitative strategy is nearly twice as frequent as the placeholder strategy (Vallejos Yopán 2023: 665). The proportions of placeholders carrying nominal and verbal morphology are roughly equal, accounting for ~45% and ~48% of the placeholder tokens, respectively. Not all of the placeholders are followed by their target: ~30% of the nominal placeholders and a surprising 50% of the verbal placeholders lack an overt target. This is similar to what Billings & McDonnell 2025 [this volume] find for the placeholder in the Austronesian language Nasal.

4.1.2 Position

The filler occurs mostly in medial position in the utterance (approximately 77% of hesitatives and 83% of placeholders), with hardly any occurrences in initial position or by themselves (Figure 9). In initial position there are twice as many hesitatives as placeholders (5% vs. 2.3%, respectively), which might be an indication that at the beginning of an utterance speakers are still planning the overall thrust of their utterance rather than searching for a particular lexeme.

¹¹Five tokens of nominally used *ujun* carry the Similative suffix *-gatcin*, and one each carry the Decessive, Limitative, Proprietive, and Privative suffix, while four tokens each of verbally used *ujun* carry the markers of Habitual and Inchoative aspect, and one carries the Refactive suffix *-dgi*, which expresses a meaning of ‘do again’ or ‘go back’.

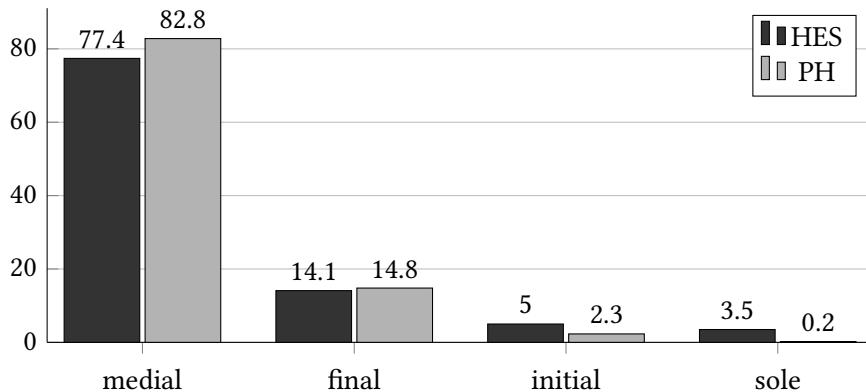


Figure 9: Position of the filler in the utterance. Unclear and indeterminate examples are excluded.

In the overwhelming majority of instances (~80%) when the placeholder is repaired, the target follows immediately after the filler, although some material can intervene. This can occasionally be an explicit word search, leading to quite long distances between the placeholder and its target. The intervening material can be recycled material (e.g. the numeral ‘one’ in (12)), a prompt by an interlocutor (see (48) in §5.1 and (58) in §5.4), or other parts of the placeholder phrase or the clause (13).

- (12) *əmən* (0) ***ujun-ma*** (550) *əmən giftca:n-ma βa:-tca:-n gun-ə-n...*
 one FLR-ACC one roe.deer-ACC kill-PST-3SG say-NFUT-3SG
 ‘He killed a **whatchamacallit**, a deer, he said, ...’ [APN_grindstone: 93; ID 392]
- (13) *babu/ka smotri taj taj gun-ə-m bi* (620) ***ujunə-l*** (0)
 granny.R look.IMP.R DIST DIST say-NFUT-1SG 1SG FLR-PL
bi-si itce-kəl kresti-l
 COP-NFUT[3PL] see-IMP.SG cross.R-PL
 “Granny, look there”, I say, “there are **whatchamacallits**, look, crosses!”
 [APN_cheremsha_brodjaga: 32; ID 300]

4.1.3 Mirroring on placeholders and recycling with targets

Nearly 75% of the placeholders with an overtly expressed target fully mirror the morphology of the target, with slightly more nominal placeholders (77%) mirroring the target morphology fully as compared to verbal placeholders (72%).

These proportions of full mirroring are considerably higher than those found in the polysynthetic Australian language Dalabon, where only 30-50% of nominal placeholders and 5-10% of verbal placeholders fully mirror the morphology of the target (Ponsonnet 2025 [this volume]). As can be seen in Figure 10,¹² the proportion of placeholders carrying only partially mirrored verbal morphology is twice as high as that with partially mirrored nominal morphology (24% vs. 12%). In Evenki, too, verbal placeholders show a higher proportion of partial mirroring than nominal placeholders (Klyachko 2022: 211). As suggested by Françoise Rose and an anonymous reviewer, this higher frequency of partial mirroring with verbal placeholders might be linked to the fact that there are more verbal derivational suffixes than nominal ones, and it is the derivational suffixes that tend to be omitted from the placeholder, as detailed below.

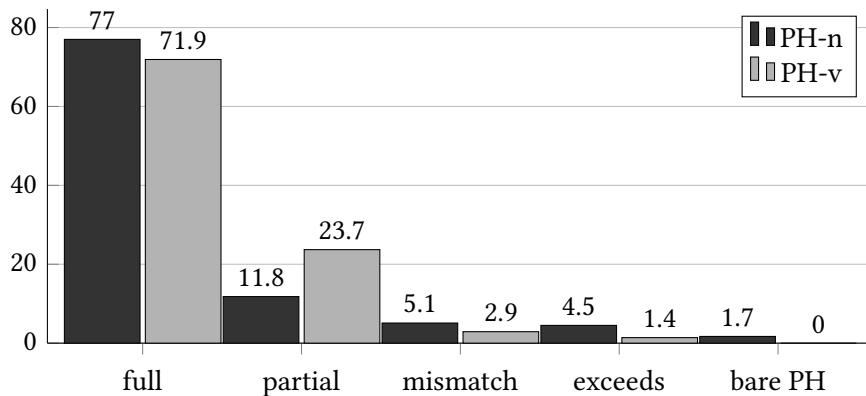


Figure 10: Mirroring of target morphology on placeholders that stand in for nouns and NPs (PH-n) and verbs and VPs (PH-v) in per cent of occurrences with targets carrying “mirrorable” morphology. Instances where there is no overt target or where the placeholder and target are in Nominative case (which in Negidal is unmarked) were excluded.

The verbal placeholder does not carry any valency-changing morphology, such as the Causative suffix *-βkan* or the suffix *-β* that can change a transitive verb into an intransitive one (14) and an intransitive verb into a transitive one (cf.

¹²The label “partial” designates instances when the PH carries less morphology than the target and “exceeds” refers to instances when it carries more morphology than the target. The code “mismatch” was used for instances in which both PH and target carry morphology, but this is not the same (this can include occurrences where there is also a mismatch in the number of morphemes), while “bare PH” designates unmarked *upun* standing in for a target that carries morphology.

Aralova & Pakendorf 2022: 29–31), even when the target carries such morphology. This indicates that in its verbal reading *uŋun* functions as an ambitransitive verb (cf. §4.1.5).

- (14) *ta-du* *əmən se:n-i-n* (2350) ***uŋun-ə-n*** (0)
DIST-DAT.ESS one ear-3SG FLR-NFUT-3SG
loku-tce-β-βə-n *ge:-n* *se:n-βa-n* (860) ***uŋun-tca:-l***
hang.up-RES-VAL-NFUT-3SG other-3SG ear-ACC-3SG FLR-PST-PL
(3660) *do:-ski* *i:-βkan-tca:-l*
inside-ADV.BALL enter-CAUS-PST-PL
‘Then one ear (of the hat) whatchamacallits, hangs, and they **did**
whatchamacallit, they tucked the other one up inside.’

[GIK_olan: 7; ID 769-770]

It is generally inflectional morphology that is mirrored: non-finite converb suffixes and the Negative Converb (15), or finite tense¹³ and subject markers for verbal placeholders, and case (16), possessive, or number marking for nominal placeholders.

- (15) *taj digi-la-βa* (0) ***uŋun-mi*** (1310) ***nupki-mi***
DIST four-NUM(DAYS)-ACC FLR-SS.COND to.smoke[TR]-SS.COND
ela-la-βa *nupki-mi* *o-ŋati-nen* *e:ma=da*
three-NUM(DAYS)-ACC to.smoke[TR]-SS.COND NEG-DEONT-3SG which=ADD
dilkān (680) ***uŋun-a*** (590) *do:-ja* *olo-du*
fly FLR-NEG.CVB to.land-NEG.CVB fish-DAT.ESS
‘If you **whatchamacallit**, if you smoke it for four days, if you smoke it for
three days, not a single fly will **whatchamacallit**, will land on the fish.’

[DIN_jukola: 97; ID 618-619]

- (16) *e:da si bi* (0) ***uŋun-mə-β*** (740) *jebga:-βə-β* *jep-tca:-s*
why 2SG 1SG FLR-ACC-1SG food.EVK-ACC-1SG eat-PST-2SG
‘Why did you eat **my whatchamacallit**, my food?’ [APK_fox: 153; ID 193]

In partial mirroring it is by and large derivational morphology that is omitted, such as the Multiplicative (17) or the Resultative ((14) above) missing from verbal placeholders, or the Diminutive suffix (18) that is omitted from nominal

¹³Note that I counted two examples where the placeholder carried the Future suffix *-ja* and the target carried the suffix *-jiŋa* as being fully mirrored, since these Future suffixes seem to occur in free variation in Negidal.

placeholders. Only four examples of verbal placeholders involve mirroring of aspectual morphemes, out of a total of 26 examples where the target carries such a suffix; no other derivational morphology (such as the Associated Motion suffix or the Multiplicative or Semelfactive) was found on the filler, and none of the seven evaluative morphemes found on overt nominal targets was mirrored on the placeholder. This is again similar to Evenki, where derivational morphology is omitted from the verbal placeholder (Klyachko 2022: 211). Thus, the placeholder carries the inflectional morphology that is needed for syntactic integration, while morphemes that do not play a syntactic role, such as evaluatives or aspectual markers, are omitted. The filler can therefore be said to be substituting for the stem rather than the root of the target.

- (17) *a taj doldi-tca:ki-nin taj bəjə oj uyun-ma bipag-ga*
and.R DIST hear-REM.PST-3SG DIST person PROX FLR-ACC settlement-ACC
hul-nakan (0) ***uyunə-ji-βa-n*** (180) ***pəktəru-ktə-ji-βa-n***
go-SS.SIM FLR-PRS.PTCP-ACC-3SG shoot-MULT-PRS.PTCP-ACC-3SG
ŋa:lə-l-tca
to.fear-INCH-PST[3SG]
'And he heard that the man walking through the whatchamacallit, the
village **is doing whatchamacallit, is shooting around**, he got frightened.'
[DIN_rok: 46; ID 685]

- (18) *jaβa-ja-n* (1580) ***uyun-ma*** (360) ***kala-ka-tçan-ma...***
grab-NFUT-3SG FLR-ACC pot.EVK-DIM-DIM-ACC
'He takes a **whatchamacallit, a little pot...**' [AET_bear: 27; ID 5]

Recycling denotes the repetition of material that precedes the placeholder with the target (Podlesskaya 2010: 23–25). This is quite rare in the Negidal corpus: only in about one-fifth of the examples with recyclable material in the placeholder phrase is this repeated with the target. In particular, adjuncts, direct objects, nominal modifiers (47), demonstratives (24), and the negative auxiliary (15) tend not to be recycled with the target. In contrast, verbs governing direct objects can be omitted or recycled with equal frequency; thus, in (19) the verb *essaβun* ‘we reach’ is not repeated with the target, whereas in (20) the verb *issen* ‘he plucks’ is recycled. In some instances there is recycling, but with a change in form, such as a Proximal Demonstrative being changed to a Distal Demonstrative. In the examples, the curly brackets frame the placeholder phrases containing recyclable and recycled material, where present.

- (19) *ti=tti {es-sa-βun be... (0) **uŋun-ma}** (0) mugdin-ma*
 like.this=PTL reach-NFUT-1PL.EX FS FLR-ACC riverbank-ACC
 'Like this we reached whatchamacallit, the bank.' [DIN_zabludilisj: 26; ID 720]
- (20) *təpu-kə-la:-jə-n tceβkan tik-kə-n gə*
 knock.out?-SUDDEN-SMLF-NFUT-3SG bird(small) fall-NFUT-3SG DP
*{is-se-n (820) **uŋun-ma-n**} (1020) {is-se-n*
 pluck-NFUT-3SG FLR-ACC-3SG pluck-NFUT-3SG
ləpultə-βa-n}
 feathers-ACC-3SG
 'He knocked [it] out, Chevkan fell down, he plucks whatchamacallit, he
 plucks the feathers.' [APN_skazka_ptica: 22; ID 457]

4.1.4 The filler as a nominal placeholder

As a nominal placeholder, *uŋun* can substitute for proper (21), higher animate (22), lower animate (23), and inanimate nouns (24).

- (21) *oj=da (160) **uŋun-ŋasa** (0) ba:tur-ka-ŋasa=da*
 PROX=ADD FLR-DEC PN-DIM-DEC=ADD
əmə-β-βat(-tce-n)
 come-VAL-HAB-NFUT-3SG
 'And this one too, late what's-his-name, late Batur, he used to bring.'
 [APN_DIN_conversation: 39; ID 318]
- (22) *taj boa baja-sal bijagi-tin boagda-du-n*
 DIST FS rich-PL.HUM settlement-3PL yard-DAT.ESS-3SG
*itce-t-tce-n itce-je-n (1050) **uŋun-ma** (1330) olgin-ma*
 see-TAM2-NFUT-3SG see-NFUT-3SG FLR-ACC pig-ACC
 'In the settlement of those rich people he saw a whatchamacallit, a pig in
 the yard.' [TIN_swine: 61; ID 879]
- (23) *gə taj kuhi-rə (0) **taj uŋu-l-ji** (610) taj*
 DP DIST to.fight.EVK-NFUT.EVK[3PL] DIST FLR-PL-INS DIST
dilka-l-ji
 fly-PL-INS
 'So they are fighting with those whatchamacallits, those flies.'
- [APN_koldekan: 94; ID 429]

The filler can also stand in for different types of NPs (cf. Schegloff 1979: 263; this is also found in the Tungusic language Even (Matić 2008) and in the Nakh-Daghestanian languages Udi and Aghul (Ganenkov et al. 2010: 101)): attributive, conjoined (25), possessive, and NPs with a relational noun functioning as a post-position (26). The attributive NPs can be quite elaborate and include relative clauses, which in Negidal are participial (27).

- (25) *je taj hətəka:si-l-tca:-du-n taj uj*
INTERJ DIST jump[ITER]-INCH-PST.PTCP-DAT.ESS-3SG DIST recently
geda-tca:ki-nin (0) uyun (0) tceβka:-tca-l siŋəjə-tca-l je
shove.into-REM.PTCP-3SG FLR bird-DIM-PL mouse-DIM-PL INTERJ
uyun-duki-n tət̪çanŋa-duki-n hətəkən-a
FLR-ABL-3SG clothing-ABL-3SG jump[SMLF]-NFUT[3PL]
‘When he started jumping, those whatchamacallit, little birds and mice,
which he had hidden, jumped out of whatchamacallit, out of his clothes.’

It can also substitute for adjectives in predicative position (28); however, only two such examples occur in the corpus. Note that the Evenki filler *aŋi/aŋa* can also stand in for predicative adjectives, as shown by an example provided by Klyachko (2022: 207, her ex. (12)).

- (28) *nɔŋan bi-tca-n so* (0) ***uŋun*** (340) *ujimkun i mu:-du jaʃu-tca-ja-n*
3SG COP-PST-3SG very FLR light and.R water-DAT.ESS
take-RES-NFUT-3SG

‘It was very whatchamacallit, light and floated upon the water.’

[AVK_crossing_stream: 105; ID 492]

The target is often provided subsequently (see for instance (25)-(28)), but this need not be the case – as mentioned in §4.1.1, the target is left unspecified in nearly 30% of the examples with a nominal placeholder. Such omission frequently occurs because the target is obvious from the speech situation, as in (29) taken from a procedural text about hide preparation: this was uttered at the moment when the speaker turned to pick up a particular tool. Occasionally, the target is described with a clause, such as the headless relative clause in (30). Note that in this example the speaker finally resorts to using the Russian lexeme that she seems to have wanted to avoid, since she cannot come up with a Negidal equivalent. There is no indication that *uŋun* can have a vague, generic meaning rather than standing in for a particular noun or noun phrase that is eluding the speaker.

- (29) *jaʃa-m ge:* (590) ***uŋun-ma*** (590)
take[NFUT]-1SG other FLR-ACC

‘Then I take the next whatchamacallit (instrument).’

[DIN_preparing_hide: 157; ID 661]

- (30) *ulguma-ja-n min-du na:-ʃɔ-m taj aŋun... (0) uŋun-mi do:-du-n* (570) *taj olo-ŋi-n na:-ktci-ji-n*
ask-NFUT-3SG 1SG.OBL-DAT.ESS put-Q.FUT-1SG DIST FS FLR-PRFL.SG
inside-DAT.ESS-3SG DIST fish-POSS-3SG put-DUR-PRS.PTCP-3SG
do:-du-n (830) ***uŋun-du-j*** (470) *kul-du-j*
inside-DAT.ESS-3SG FLR-DAT.ESS-PRFL.SG sack-DAT.ESS-PRFL.SG
‘She asks me whether to put it into the whatchamacallit (lit. She asks me “should I put it into my whatchamacallit”), into that where she puts the fish, in the whatchamacallit, in the sack.’ [GIK_shuka: 38-39; ID 787-788]

There are furthermore three examples in which the filler appears to be standing in for direct speech (31), as shown by prosody and the fact that in two examples the fillers were translated as speech complements,¹⁴ indicating that our consultant interpreted them as holding the place for direct speech utterances. Furthermore, *ujun* is marked with the Accusative case expected for the complement of the speech verb in example (31). Unfortunately, two of these examples come from one of the old digitized tape recordings, so that the extraction of the pitch contour in Praat is full of errors. Nevertheless, a rise in pitch on the first syllable can clearly be identified in Praat (Figure 11); this differs from the prosody of hesitant uses of *ujun*, justifying the interpretation that *ujun* has a placeholder function here.

- (31) *minə-βə gun-tca-n* (820) *ujun-ma* (1570) *katçikan-ma-s bi jepu-βka-ŋta...*
 1SG.OBL-ACC say-PST-3SG FLR-ACC puppy-ACC-2SG 1SG
 eat-CAUS-HORT.SG

‘It said this to me: “Let me feed your puppy,...”

[APK_1chindakan: 62-63; ID 153]

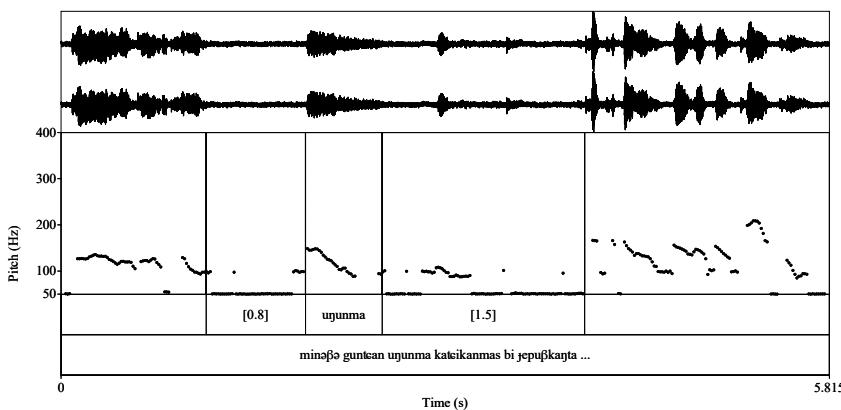


Figure 11: Waveform and pitch contour of example (31)

Such uses of the filler are also attested in Evenki, but Klyachko (2022: 207) analyses them as instances of hesitant uses rather than as placeholder.

¹⁴I reflect this in the translation of (31) by exceptionally not translating the placeholder as ‘whatchamacallit’.

4.1.5 The filler as verbal placeholder

As a verbal placeholder, *uŋun* can stand in for both intransitive (32) and transitive (33) verbs without taking any additional valency-changing morphology, similar to the Manambu placeholder verb *məgi-* (Aikhenvald 2008: 399–400, 575). This implies that *uŋun* can be analysed as an ambitransitive verb – which explains the lack of valency-changing morphology on the filler even when the target carries such morphology, as mentioned above (§4.1.3). The filler can also stand in for a VP (see the coding definitions for details on how these were identified), substituting for the verb plus adjunct (8, 14) or direct object.

- (32) *atika:-sal* *gə təgə-dgi-ja:n* (890) *unu-l-tca:-l* (0)
old.woman-PL.HUM DP sit.down-REP-SS.ANT FLR-INCH-PST-PL
məjga:-l-tca:-l *gun-ə*
think-INCH-PST-PL say-NFUT[3PL]
‘The old women, having sat down, started to whatchamacallit, started to think, they say...’ [GIK_olan: 23; ID 768]

(33) *tadukin aja ajag ajanis-sa:n* *opuška-ja-n* (0)
then good FS adjust.properly-ss.ANT trimming.R-DEST-3SG
unuŋ-ə-s (280) *tulə-s*
FLR-NFUT-2SG attach[NFUT]-2SG
‘Then, having adjusted (the centre) properly **you** whatchamacallit, you add the trimming.’ [DIN_duck_heads: 75; ID 577]

However, even though as a verb *ujun* is mostly used as a placeholder (e.g. (6), (14), (15), (17) above), both with and without an overt target, there are also several instances where it functions as a generic action verb.¹⁵ These uses probably account in part for the very high frequency (50%) of verbally used *ujun* without an overt target (§4.1.1). It can refer to a general, unspecified action ('do'), as in (34), which is taken from a narrative about a game in which the loser has to fulfill the wish of the winner. Here, the loser is asking the winner what it is he should do for her, i.e. the reference is clearly to an as yet unspecified action. The filler can also be used deictically (mostly accompanied by the manner demonstrative *ti*: 'like this'), referring anaphorically to what was said in the preceding discourse (35) or exophorically to what is being done in the speech situation (36).¹⁶ Such examples

¹⁵In §4.1.1 to §4.1.3 all tokens of *uyun* carrying verbal morphology were included in the counts for verbal placeholder.

¹⁶I refrain from adding pause lengths to this example, since the pauses were the result of the demonstration and are hence not indicative of any form of disfluency.

with exophoric meaning are mostly found in the procedural texts, and in the accompanying video one can see that the filler is accompanied by the appropriate action (Figure 12).

- (34) *gə e:kun-ma* (0) ***uŋun-ŋati-β*** (0) *gun-ə-n*
 DP what-ACC FLR-DEONT-1SG say-NFUT-3SG
 “Well, what shall I do (for you)?”, he says. [DIN_games: 34; ID 594]
- (35) *a osi ti-kan* (0) ***o-ta-s*** ***uŋun-a***
 and.R now like.this-DIM NEG-NEG.FUT-2SG FLR-NEG.CVB
 {In the past we fished as much as [we needed]. There was a little house
 like that up high [on piles] and there we put the dried fish so that it
 would last all year for us and for our dogs.} ‘But now **you can’t do** like
 that.’ [AET_village_life: 90; ID 135]
- (36) *sa:-ŋati-s* *olgo-tca-βa-n* ***ti:***
 know-DEONT-2SG dry.out[INTR]-PST.PTCP-ACC-3SG like.this
uŋun-e-ki-s ***ti:*** ***uŋun-gi-je-ki-s***
 FLR-NFUT-COND-2SG like.this FLR-REP-NFUT-COND-2SG
o:-ŋati-nin *nəptə-mdi*
 become-DEONT-3SG smooth-INTS.ADJ
 ‘You will know that it’s dry if, when **you do like this** and **when you do it again**, it straightens out.’ [DIN_preparing_hide: 185; ID 662-663]



ti: uŋunekis



ti: uŋungijekis

Figure 12: Two still frames taken from the video that illustrate the rolling up and unrolling gestures that accompany the tokens of *uŋun* in example (36)

Example (36) is taken from a recording in which the speaker explained and demonstrated how the Negidals prepare elk hide. Here, she shows how you can judge whether you have scraped off all the moisture by rolling it up ('when you do like this') and then unrolling it again ('you do it again' – with the refactive suffix *-(d)gi*).

There are also examples in which *uŋun* is used not as a placeholder for a target that eludes the speaker nor with anaphoric reference, but to substitute for a verb that was used in the immediately preceding context (37).¹⁷ Such uses of the filler have also been identified for Northern Pastaza Kichwa, where Rice 2025 [this volume] terms them 'resumptive pro-verbs'. The function of such resumptive uses of the filler is unclear, but it is unlikely to be a problem with the planning or execution of speech. Nor is it likely to be the avoidance of repetition of the lexical verb, as shown by the fact that the sentence following (37) contains three tokens of the verb 'eat'.

- (37) *taj bəjun ti: jepu-βki, jaJa ti: a-βki uŋun-a*
 DIST elk like.this eat-HAB.PTCP bear[EUPH] like.this NEG-HAB.PTCP
FLR-NEG.CVB
 'An elk eats like this, but not a bear. (Lit. An elk eats like this, a bear
 doesn't do). {Why would a bear eat a tree? A bear eats berries, it eats
 nuts.}' [APN_DIN_galigda: 210; ID 345]

In the verbal uses in which it does not function as a placeholder, *uŋun* partly overlaps with the generic action verb *neko-* 'do'. Next to its use to refer to unspecified actions (38), sometimes with anaphoric (39) or exophoric reference, *neko-* is used as a light verb with ideophones (40).

- (38) *e:kun-ma neko-ŋa:ti-s*
 what-ACC do-DEONT-2SG
 'What are you going to do?' [GIK_1belekakta: 1]
- (39) *ti: neko-ja:n mani-n ju:-ja-n ŋənə-jə-n*
 like.this do-SS.ANT self-3SG exit-NFUT-3SG go-NFUT-3SG
 'Having done this he himself exits and leaves.' [DIN_girl_devil: 77]

¹⁷This is an example without a sound file, hence the pause duration cannot be established.

- (40) ... *araj taj-du-n uyun-du-n həjə-du-n*
... RESTR.EVK DIST-DAT.ESS-3SG FLR-DAT.ESS-3SG bottom-DAT.ESS-3SG
e:kun=ka kope:r kope:r neko-jə-n
what=FOC IDEO IDEO do-NFUT-3SG
‘... only something at the bottom [of the sack] is rattling (lit. **does**
koper-koper).’ [APK_fox: 107]

The generic action verb *neko-* is furthermore frequently used as an auxiliary in analytical constructions with a prospective reading, that is, in constructions with a reading of imminence, often with a nuance of intention or desire (see Matić 2017 for a description of this construction in Negidal’s sister Even). This construction consists of the lexical verb in the Reflexive-marked Purposive Converb form *-da-j* (sg) or *-da-βaj* (PL) with *neko-* acting as auxiliary (41). There are three examples in the corpus where *uyun* takes the place of the auxiliary *neko-* in such prospective constructions, e.g. (42), but no examples where *uyun* functions as a light verb with ideophones.

- (41) *itce-t-tce siβu-ŋi-tin ələ təgə-da-j*
see-TAM2-NFUT[3PL] sun-POSS-3PL already sit.down-vs.PURP-PRFL.SG
neko-l-tca:
do-INCH-PST[3SG]
‘They look: the sun is about to go down. (lit. is about to sit down)’
[GIK_kljukva: 16]

- (42) *puktəran-da-j ələ (0) uyun-tca-n, (0) ...*
shoot-vs.PURP-PRFL.SG already FLR-PST-3SG
‘He was already about to shoot...’ [AET_bear: 20; ID 4]

However, while the filler overlaps partly with *neko-* in its use as a generic verb, *uyun* is the only verbal placeholder: there are no examples in the corpus in which *neko-* substitutes for another verb.

In three examples in the corpus, which were all uttered by the same speaker, the filler seems to be substituting not for a verb (phrase), but for an independent clause. In two examples, one of which is shown here (43), it is a 3PL Non-Future tense-marked filler that refers to the following clause, as shown by the prosody, which is clearly that of a syntactically integrated verb and not a hesitative (Figure 13), and the translation provided by our consultant, which included the filler.

- (43) *gə taj bi-je-mne:n* (0) *uŋun-ə* (860)
 DP DIST live-DUR-SS.DUR FLR-NFUT[3PL]
əma:tcin-a-βaj *o:-na:-tca-l* *di:-ski*
 birchbark.boat-DEST-PRFL.PL make-AM-PST-PL taiga-ADV.B.ALL
 ‘Well, living like that they do something, they went to the forest to make
a boat for themselves.’ [APN_two_sisters: 4-5; ID 485]

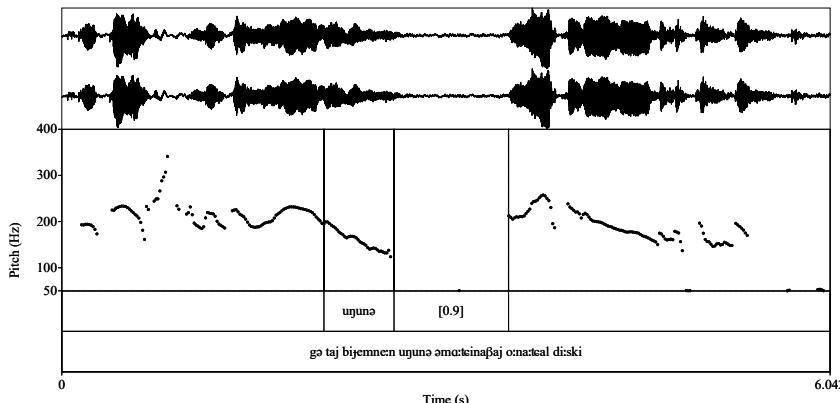


Figure 13: Waveform and pitch contour of example (43)

In these examples, *uŋun* is translated with ‘do (something)’ (что-то делают), similar to its use with anaphoric or exophoric reference, but here the reference is cataphoric, to an event that has not yet been mentioned by the speaker. Whether this has a particular discourse function or is merely a means to gain time to gather one’s thoughts cannot be judged given the small number of examples.

In the third example (44), the filler is unmarked, yet prosodically it can be identified as a placeholder, since its pitch does not drop down to the level often found for hesitations (Figure 14). Furthermore, it was translated as the complement of ‘see’ by our consultant. The substituted clause is finite and thus grammatically not a complement clause (which in Negidal would take an Accusative-marked participial predicate), but the context makes it clear that what the speaker sees is the boat coming.

- (44) *itce-je-βun* (0) *uŋun* (0) *əmən onotco-kan* *əmə-ja-n*
 see-NFUT-1PL.EX FLR one boat.EVK-DIM COME-NFUT-3SG
 ‘{We went there. Going we saw.} We saw **something**, a boat is coming.’
 [APN_anecdotes: 16; ID 275]

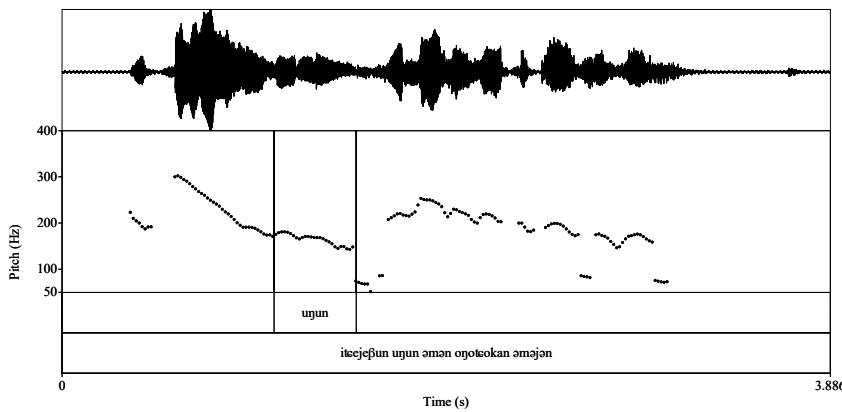


Figure 14: Waveform and pitch contour of example (44)

To summarize what was said in §4.1, the Negidal filler *uŋun* occurs with high frequency in spoken discourse, with the placeholder strategy dominating over the hesitative strategy; however, targets are frequently left unexpressed. It takes the place of nouns and verbs with equal frequency and thus goes against the cross-linguistic tendency for placeholders to occur more commonly with nominal than with verbal targets (Podlesskaya 2010: 13). While the filler carrying verbal morphology often has a reading of a generic action verb, *uŋun* used as a noun does not seem to have any vague or generic meaning. In addition to substituting for nouns and verbs the filler can also stand in for clauses and direct speech, although this is very rare in the data and restricted to the two speakers who were born before 1925. This is not the only inter-speaker difference in use of the filler, as will be addressed in the following section.

4.2 Differences in use of *uŋun* between speakers

As can be seen in Figure 15, the frequency of the filler varies considerably across speakers. Such inter-speaker variability has also been found for markers of disfluency in German women (Braun et al. 2023), and for the use of placeholders in Kalamang (Visser 2025 [this volume]) and Besemah (McDonnell & Billings 2025 [this volume]) spoken in Indonesia, Kolyma Yukaghir spoken in Siberia (Ventayol-Boada 2025 [this volume]), and among three speakers of the endangered Australian language Dalabon (Ponsonnet 2025 [this volume]).

Interestingly, there is no linear correlation with the proficiency of the speakers (Figure 15). It is likely that the two speakers who use the filler most, namely

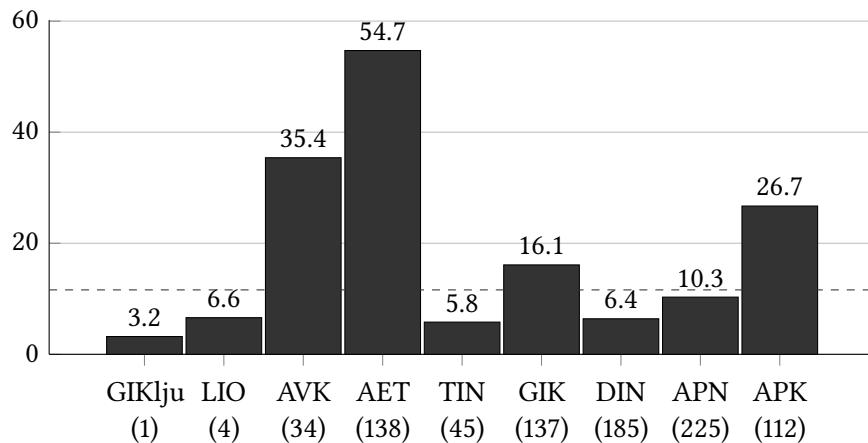


Figure 15: Frequency of use of *uŋun* by speaker, normalized to 1000 words. Speaker ID's are arranged in order of increasing proficiency, and the absolute number of *uŋun* they produced is indicated in brackets. The dashed line indicates the corpus average (11.6 per thousand words).

AVK and AET, do so because they are relatively weak and insecure speakers (Pakendorf & Aralova 2018). The influence of language proficiency on the rate of filler use has also been demonstrated for Evenki (Klyachko 2022: 213–214) and been observed for filled pauses in general (Kosmala & Crible 2022: 220). In this context it is a bit surprising that the two weakest speakers, GIKlju and LIO, produce hardly any *uŋun*. However, for her own contribution to the corpus GIKlju only read a prepared text, which not unsurprisingly contains no fillers; the one filler she did produce was while trying to ask AVK a question in Negidal during the recording of the latter's narrative. With respect to LIO, it might be that she was actually more proficient in Negidal than the few very hesitant narratives recorded with her led us to believe – this is indicated by a fluent spontaneous interaction with her mother which was recorded when she served as an interlocutor. Rather harder to explain is the relatively high frequency of use of the filler by the only male speaker, APK, who is said to have had an “excellent knowledge of the language” (Khasanova & Pevnov 2003: 227, my translation). Possibly his above-average use of the filler is due to a certain pressure to narrate folktales in a specific way using particular terms and phrases (Khasanova & Pevnov 2003: 233–234). This might have induced him to hesitate frequently (see Figure 17 below) while he remembered necessary details.

An alternative explanation for the inter-speaker variability in frequency of filler use, suggested by an anonymous reviewer, is the impact of differences in

Negidal practice among different speakers. Since Negidal is hardly used in everyday life anymore, the speakers who contributed to the corpus would have gained fluency over the period they worked with the recording linguists. Speakers who did not contribute much text would have remained relatively rusty in Negidal, and hence resorted to the filler a lot while searching for forgotten words, while speakers who produced more recordings would have acquired practice speaking it for the linguists and hence produced less fillers overall. However, a breakdown of the frequency of filler use by year for the three speakers (APN, DIN, and GIK) for which the corpus contains recordings made over several years (Figure 16) does not show the linear decrease over time that would be expected if language practice were the only factor at play. It is true that there is a drastic drop in occurrence of the filler in GIK's production between 2005 and 2006, which might indicate some form of habituation. However, based on my own work with GIK, I think it is more likely diffidence rather than rustiness that made her hesitate, with the drop in filler use in later years being due to her increasing familiarity with the recording linguists.

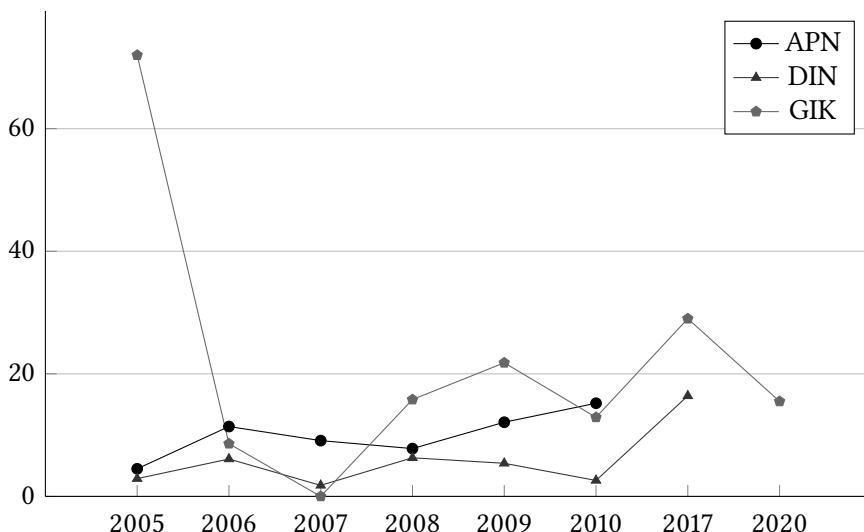


Figure 16: Frequency of filler use (per 1000 words) by year of recording for the speakers APN, DIN, and GIK. These are the only speakers for whom the corpus contains recordings made over several years as documented by metadata.

Speakers do not only differ in the frequency with which they use fillers, but also in their choice of hesitatives vs. placeholders (cf. Visser 2025 [this volume])

for a similar finding for the Papuan language Kalamang, albeit taking into account also the hesitative ‘uh(m)’, which is excluded in my study). DIN makes very little use of *ujun* as a hesitative, with only 10% of her tokens falling into this category (Figure 17); this contrasts strikingly with AET, nearly half of whose tokens of the filler (46%) are hesitatives.

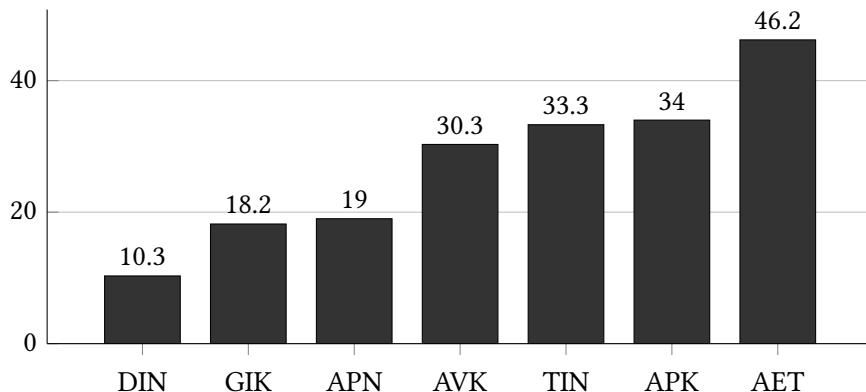


Figure 17: Proportion (in per cent) of hesitative use of filler tokens produced by individual speakers, arranged by increasing frequency of hesitatives. Unclear and indeterminate tokens were excluded from the total count. Since GIKlu and LIO produced hardly any fillers, they are omitted from the figure.

This difference – which results in a difference between fluent vs. hesitant speech – can probably be partly explained by differences in language proficiency, since DIN, GIK, and APN are relatively proficient speakers, while AET is one of the speakers with the least active use of Negidal and hence lacks proficiency. In addition, the pressure to produce narratives in a particular way and to speak “proper” Negidal might also have an impact on the amount of hesitative vs. placeholder uses of the filler. Thus, as mentioned above, the fact that APK, a proficient speaker, uses the filler as a hesitative rather than placeholder in 35% of his tokens might be due to the pressure to narrate fairy tales in a precise way following particular conventions. This could have increased his hesitancy without increasing the need to search for lexemes. Similar pressures can explain the relatively high frequency of hesitative uses for TIN (33%) and AVK (30%): as mentioned in §2, the former recorded little narratives for use as Negidal lessons to be broadcast over the local radio, which would have increased the pressure to use only Negidal words. As to AVK, she was very self-conscious and tried hard to speak “correct” and “pure” Negidal in her contribution to the archive – to such an extent that she

discarded a first recording of her narrative, which she deemed too disfluent, and asked to be recorded again in the presence of a Negidal interlocutor (GIKlju). In contrast, the recordings involving DIN and APN, in particular, were often very informal and even include conversations at the tea table, which would have lowered the pressure to produce “proper” Negidal; hence, they use the filler mostly when searching for particular words or in its function as a generic verb.

Furthermore, DIN stands out with an extremely high frequency of verbal place-holders, twice as high as her use of nominal placeholders (Figure 18). This might be due to the fact that she produced several procedural texts, in which *uŋun-* is frequently used anaphorically and exophorically, as discussed in §4.1.5. DIN’s sister GIK, too, produces the verbal placeholder relatively frequently, in contrast to APK, TIN, and APN, in whose productions it is the nominal placeholder that predominates.

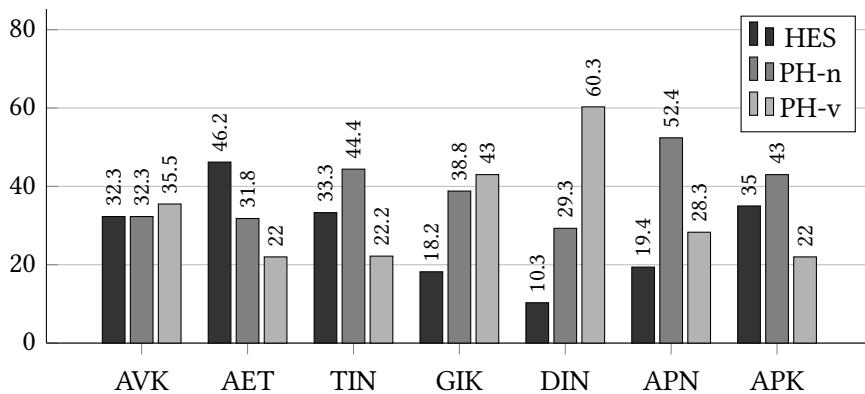


Figure 18: Frequency of strategies of use of *uŋun-* by individual speakers (arranged in order of increasing proficiency), excluding unclear and indeterminate tokens as well as those that stand in for adjectives, clauses, and direct speech. Since GIKlju and LIO produced hardly any fillers, they are omitted from the figure.

It is notable that with the exception of AVK all the speakers omit the target more often for verbal placeholders than for nominal placeholders (Figure 19). As suggested by Françoise Rose, this could be due to the fact that the verbal placeholder has a generic meaning of ‘do’ (cf. §4.1.5), while the nominal placeholder lacks a similar generic meaning (§4.1.4). This is particularly noticeable for APN, who omits the target in over 70% of her uses of *uŋun-* as a verbal placeholder, leaving it up to the hearer to reconstruct the meaning. APN also omits nominal targets very frequently (nearly 40% of all her tokens of the filler standing in for

a noun or NP lack an overt target). Possibly her advanced age – she was in her early nineties when she was recorded – made her unwilling to expend the effort needed to search for words that eluded her.

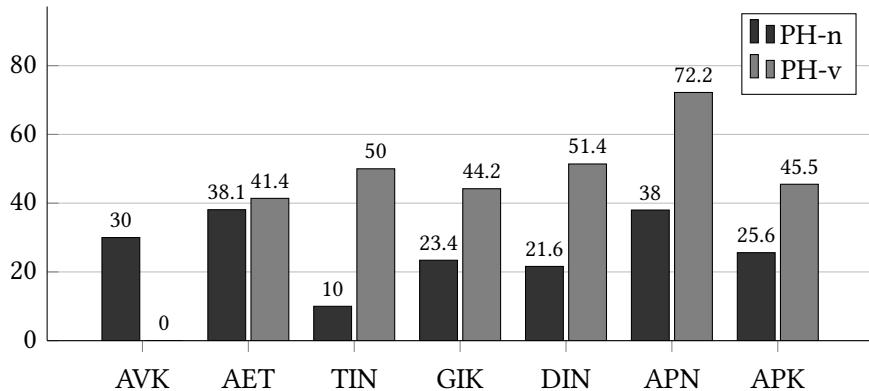


Figure 19: Proportion of placeholder uses without overt targets across speakers. GIKju and LIO are omitted, since the extremely small number of fillers that they produced precludes meaningful analysis.

To summarize, speakers resort to using the filler with very variable frequency, and they also vary in the frequency with which they use it as a placeholder or hesitative. The choices seem to be governed by proficiency of the speakers as well as by the genre and recording context. Less proficient speakers use the filler more frequently, and in particular as a hesitative rather than as a placeholder; similarly, the pressure to narrate folk tales in a particular manner produces many fillers, especially hesitatives. In contrast, procedural texts trigger the use of targetless verbal placeholders with a generic reading.

5 Cognitive and pragmatic aspects of filler use

5.1 Fillers in word search

As mentioned in the introduction, fillers are one of the strategies speakers can use to play for time when they encounter difficulties in retrieving a lexical item. This is demonstrated particularly clearly by examples where the word search is subsequently verbalized (45), but also by instances where a placeholder is broken off when the speaker remembers the target (46):

- (45) *na:n dun man mo:-ŋi-n bi-si-n taj (0) uŋun* (1450) *uŋun*
 3SG.DAT.ESS self tree-POSS-3SG COP-NFUT-3SG DIST FLR FLR
 (1890) *bajə mo:-ŋi-nen taj... listvennica omjo-tca bi-si-m*
 man tree-POSS-3SG DIST larch.R forget-PST.PTCP AUX-NFUT-3SG
 ‘He has his own tree [i.e. for men there is a separate tree on which to
 perform the holgin ritual], that **whatchamacallit/uhm...**
whatchamacallit/uhm... A man’s tree is.... I forgot (how to say) larch
 [using the Russian word for ‘larch’].’ [GIK_ostroe 20-22; ID 776-777]
- (46) *taj=tii bu jeli-βa (1000) uŋun-e-β... (0)*
 DIST=PTL1 1PL.EX taimen-ACC FLR-NFUT-1
olo-ma-t-tce-βun
 fish-VR-TAM2-NFUT-1PL.EX
 ‘Like that we **whatchamacall...**, caught a taimen!’ [DIN_rybalka: 27; ID 691]

Very often, when the delayed constituent is finally produced, it carries a higher pitch which emphasises it (cf. Podlesskaya & Korotaev 2022: 71–73, for a similar observation in Russian). This kind of emphatic pitch can follow after both hesitative uses of *uŋun* and placeholder uses: in example (47), the first token of the filler is not morphosyntactically integrated and is hence identifiable as a hesitative, with the delayed constituent *bakaja* ‘they found’ pronounced with emphatic high pitch on the first syllable (Figure 20). The second token of the filler carries Accusative case and is clearly the placeholder for the NP *tcoltcoki sama:nman* ‘the shaman of the weasels’; this carries even higher pitch on the first syllable of *tcoltcoki* ‘weasel’. This high pitch and intensity on the delayed constituent underlines the fact for the hearer that the word search was successfully concluded.

- (47) *sama:n-a gəla:ktə-na:-ja:n o:-mi (0) uŋun* (420)
 shaman-ACC.INDF look.for-AM-SS.ANT make.SS.COND FLR
baka-ja əmən (0) uŋun-ma (660) *tcoltcoki*
 find-NFUT[3PL] one FLR-ACC Siberian.weasel
sama:n-ma-n
 shaman-ACC-3SG
 ‘Having gone to look for a shaman... **uhm...** they found a
whatchamacallit, a shaman of the Siberian weasels.’
 [APK_frog_tale: 120; ID 224-225]

The Negidal data partly contradict the common assumption that speakers encounter more difficulties in retrieving infrequent lexemes or items that are less

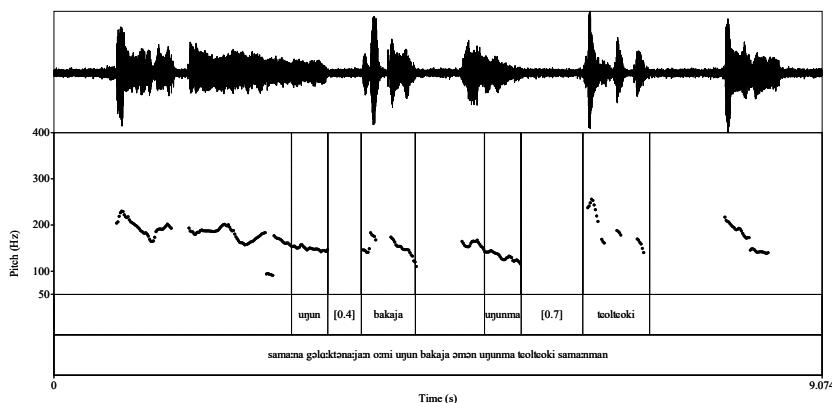


Figure 20: Waveform and pitch contour of example (47)

predictable in a given discourse context (Lickley 2015: 461–462): an admittedly rather limited analysis of the nature of the overt targets of placeholders shows that only about 40% are items that might be considered rare,¹⁸ such as names of people who lived a long time ago and whom one has hence not referred to in a while, or archaic cultural objects such as shaman’s tambourines, particular types of footwear, or the prow of birchbark boats. In contrast, nearly 39% of the fillers are triggered by common or even very common items, sometimes on multiple occasions. For instance, ‘river/stream’, which refers to a crucial feature of the Negidal environment, occurs in the corpus 63 times and triggers a filler four times, and ‘dish/container’ (55 occurrences in the corpus) elicits a filler five times. Even some very frequent verbs, such as ‘make’ (375 occurrences in the corpus) or ‘think’ (225 occurrences in the corpus) elicit three and four fillers, respectively.

There are furthermore several instances when items that were produced in the immediately preceding context trigger the use of a placeholder, as described in §4.1.5 for the ‘resumptive pro-verb’ use of *ujun*. Such rather unexpected uses of the filler have also been noted for the Maliseet-Passamaquoddy “noun substitute” by LeSourd (2003: 142):

¹⁸This was estimated by checking the frequency of items in the corpus via concordance searches, but also by assuming that the names of characters in folk tales or of people who lived a long time ago would not be commonly used and hence would no longer be activated (cf. Schegloff 1979: 274), or that the terms for archaic cultural items would be rarely used. Conversely, culturally salient items, like boats, fish hooks and lures, or the lean-to which serves as a vital storage room in houses, were considered to be common, even if they didn’t occur very frequently in the corpus. Note that when the probable delayed constituent of hesitative or indeterminate uses is included in the analysis, the relative proportions of rare vs. (very) common items remain the same.

[M]any occurrences of the noun substitute do not appear to be of this character [a hesitation signal]. Often the [placeholder] pronoun is used to introduce a common word that one might expect a speaker to find easy to recall. A speaker will sometimes employ the noun substitute to introduce a noun that he or she has already used several times in a given discourse. Not surprisingly, there is often no appreciable pause in such cases before the speaker provides the target noun. We are accordingly faced with a puzzle. Why do speakers apparently use a hesitation form even when they seem to feel no need to hesitate?

For example, in the sequence in (48) DIN asks her mother APN whether she remembers Tanja Trofimova's father, but then repeats her question using the placeholder. Apart from 'father' being a very common word (243 instances in the corpus) that one might expect to have been easily retrievable, it was produced a mere six seconds before the placeholder, and thus might be assumed to have still been activated when DIN rephrased her question – yet it is notable that she pauses for a considerable stretch of time before the placeholder.

- (48) a. DIN: *si: opi: tanja-ŋasa trofimov-ŋasa amin-ŋasa-βa-n*
2SG mother.VOC PN-DEC PN-DEC father-DEC-ACC-3SG
jo-ŋtci-s
remember-DUR[NFUT]-2SG
'Mama, do you remember the late father of late Tanja Trofimova?'
b. APN: *ni:?*
'Who ?'
c. DIN: *tanja-ŋasa trofimov-ŋasa* (1530) ***uŋun-ma-n*** { APN: *ami-nin* }
PN-DEC PN-DEC FLR-ACC-3SG father-3SG
DIN: *amin-ma-n jo-ŋtci-s*
father-ACC-3SG remember-DUR[NFUT]-2SG
'DIN: Late Tanja Trofimova's whatchamacallit... {APN: her father}
DIN: ... do you remember her father?'

[APN_DIN_conversation: 59-63; ID 320]

In some instances, it is likely that the speakers were distracted (cf. Lickley 2015: 463) and hence couldn't immediately retrieve the item they needed. Unfortunately, since we lack video recordings for most of the Negidal data, it is impossible to identify external distractors, such as a bird flying by outside or a sound at the door. Nevertheless, possible causes of distraction can be identified

in the data, as will be outlined in the following. These comprise 1) the presence of as-yet-unfamiliar linguists with a video camera and other recording devices, 2) the realization that one has made an error, or 3) the search for a different lexeme that is going on in the background.

Concerning the first point, the verb *uli-* ‘sew’, which appears in the corpus 110 times, elicited a placeholder on four occasions, once in an explanation of how to make thread out of elk sinew in which sewing was mentioned ten times. All tokens of the placeholder for ‘sew’ occurred in recordings done by linguists who were not yet very well known to the speakers and that involved both video and audiorecording. It is thus possible that the video camera and big audiorecorder in combination with foreign linguists facing the speakers (on one occasion, three linguists were present at the same time) distracted the speakers and led to their lack of concentration.

As to the second type of potential distractor, it is likely that the filler in (49) was not triggered by difficulties in retrieving the word for ‘horse’, but rather by the speaker’s realization that she made a mistake (there were three horses, and not two); this may have distracted her and led to the disfluency. This is indicated by the false start on *mo[jenma]* followed by a filled pause [əmm], and the fact that the filler, while carrying the appropriate Accusative suffix needed to substitute for ‘horse’, is pronounced with the low pitch and intensity found on the filled pause (Figure 21). In addition, there is a long pause before the numeral *ful* ‘two’, further underlining the hesitation of the speaker.

- (49) *alaβ-kal* (1930) *ful* (770) *mo... əmm uŋun-ma* (590) *elan*
 to.harness-IMP.SG two FS ehm FLR-ACC three
mojen-ma *gun-ə-n*
 horse-ACC say-NFUT-3SG
 “Harness two... whatchamacallit, three horses”, she said.
 [TIN_swine: 91; ID 882]

Lastly, as mentioned above, speakers might be distracted by a word search that is going on in the background. For example, in (50) the speaker, GIK, hesitates and resorts to the placeholder before producing what appears to be the target both on the basis of syntax (Accusative case) and intonation (emphatic high pitch and strong intensity), namely *tomkolβa* ‘threads’ – even though the root *tomko* ‘thread’ is found in the word that directly precedes the placeholder. In this case it seems as if what GIK might be struggling with is not the word ‘thread’, but rather the word for ‘rag’, as indicated by her further hesitations and the overt admission that she had forgotten the Negidal word for ‘rag’ (Russian *loskuty*).

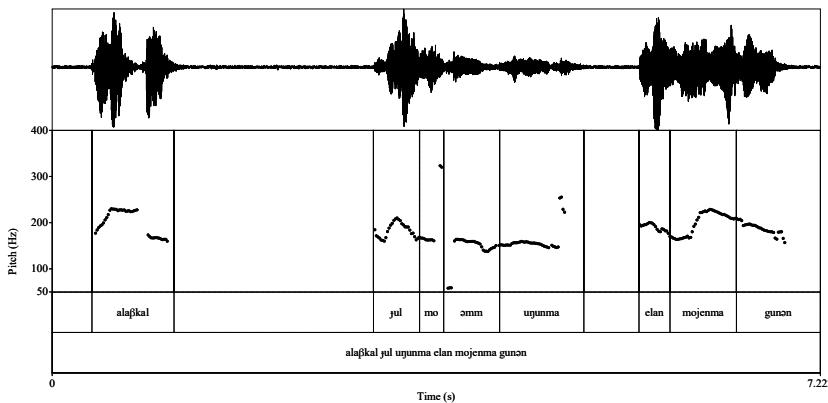


Figure 21: Waveform and pitch contour of example (49)

She may thus have been distracted by the ongoing search for this lexeme that eluded her, triggering the unexpected placeholder for the familiar and activated term ‘thread’.

- (50) *o:-ja-n jaβa-ja-n tomko-l-duki-n taj*
 make-NFUT-3SG take-NFUT-3SG thread-PL-ABL-3SG DIST
səlko-da-ββi tomko-duki-n (2550) ***ujun-ma*** (2370)
 silk.thread.R-VR-IMPERS thread-ABL-3SG FLR-ACC
tomko-l-βa (1490) *ulgə-ja-n* (750) *taduk ujun-ma*
 thread-PL-ACC to.braid-NFUT-3SG then FLR-ACC
ulgə-ja-n taj loskuty omŋut
 to.braid-NFUT-3SG DIST rags.R fs[forget]

‘He first takes some of the threads, those with which one embroiders, he braids the whatchamacallit, the threads, after that he braids the whatchamacallit, the rags - I forgot (the Negidal word).’

[GIK_ostroe: 12; ID 771-772]

Such examples indicate that the planning of an utterance is not finalized before it is begun, but continues as one speaks, and that it takes place relatively far in advance and can thus interfere with ongoing speech. They further demonstrate that it is not necessarily merely the accessibility of the target item that triggers a placeholder, but that other factors play a role as well, such as level of concentration, fatigue, physical well-being, and external and internal distractions, not all of which are easily identifiable from documentation corpora such as the one used here. Some of the factors that might have played a role in the fluctuations

in filler use over time observed for the speakers APN, DIN, and GIK (Figure 16 in §4.2) are: increasing familiarity with the recording linguists and hence decreasing formality of the recording situation (for the years 2005–2010), increasing age, level of tiredness and concentration at the moment of individual recordings, and particularities of the recording situation. For example, the steep increase in use of the filler observed for DIN in 2017 (Figure 16) might be linked to her increased age as well as to the fact that she was faced solely by unfamiliar linguists, while in the preceding years there were often several members of the family present during recording sessions.

5.2 Fillers and utterance planning

The morphological integration of placeholders (§4.1.5) demonstrates that the syntactic outline of the utterance is planned well in advance, since inflectional morphology on the placeholder, such as case on nouns or tense suffixes and subject indexes on verbs, mirror those of the target. Mismatches in the morphology found on the placeholder and the target occur very rarely, concerning only about 3% of the verbal placeholders and less than 6% of the nominal placeholders. In addition, even when there are mismatches, these often involve morphemes that are semantically congruent, e.g. a Past Tense suffix on the verbal placeholder and a Non-Future marker on the target (51), or cases with a partial overlap in functions¹⁹ for the nominal placeholder (52). In (52), the speaker takes up the bare Allative form of the target that her sister prompts her with (indicated with curly brackets), even though she had used the possessive-marked Locative case on the placeholder.

- (51) *nojaltin kogda itce-tca-tin bi ə-si-m na:n-tiki-n*
 3PL when.R see-PST-3PL 1SG NEG-NFUT-1SG 3SG-ALL-3SG
daga-ma-ja (1070) uŋun-tca-l (430) majga-ja...
 near-VR-NEG.CVB FLR-PST-PL think-NFUT[3PL]
 ‘When they saw that I am not coming to them they did whatchamacallit,
they thought...’ [GIK_1tatarskoe: 62; ID 727]

¹⁹These involve the Locative, Dative-Essive and Allative, which can all express a goal, the Accusative instead of the Prolative, since the Accusative can occasionally express a median (as in (17) above), or the use of the unpossessed Accusative on the placeholder to mark a direct object instead of the Reflexive Possessive suffix on the target, since the latter is used to mark direct objects when the possessor is coreferential with the subject of the clause.

- (52) *ŋənə-m* (570) *uŋun-dula-j* {*togo-tki*} *togo-tki* *ŋənə-m*
 go[NFUT]-1SG FLR-LOC-PRFL.SG fire-ALL fire-ALL go[NFUT]-1SG
 ‘I go to whatchamacallit. {To the fire.} I go to the fire.’
- [GIK_ojavi: 11-13; ID 762]

The difference in frequency of the hesitative vs. placeholder strategies across speakers (§4.2) indicates that individuals have different kinds of problems with speech formulation. Even though both young and old speakers (born after 1940 and before 1925, respectively) use the placeholder strategy three times as frequently as the hesitative strategy (Table 3), there is a notable discrepancy in frequency of hesitative uses of *uŋun* when comparing the less proficient speakers (AET, AVK, GIKlju, LIO) with the more proficient speakers (APK, APN, DIN, GIK, TIN). The former use the filler in its hesitative function nearly as frequently as in its placeholder function (Table 3). This is particularly pronounced for AET, as shown by Figure 17 in §4.2, and contrasts strongly with the more proficient speakers, who use *uŋun* as a placeholder four times more often than as a hesitative. This indicates that whereas the more proficient speakers mainly have recourse to *uŋun* to fill a particular slot in an already constructed syntactic frame, the less proficient speakers have problems not only with finding particular words, but with putting the syntactic frame in place.

Table 3: Comparison of hesitative vs. placeholder use by relative age and proficiency of speakers

	young	old	weak	strong
HES	24.9	24.2	42.4	19.9
PH	75.1	75.8	57.6	80.1

5.3 Fillers in repair

The Negidal filler can also be used to stall for time in order to effect a repair, as has been observed in other studies (Clark & Fox Tree 2002: 87, LeSourd 2003: 153, Ganenkov et al. 2010: 103). However, the proportion of fillers that occur in repair contexts is quite low, as has been found cross-linguistically (Lickley 2015: 463 and references therein), involving only 5.6% of all fillers.

In the current language ecology of Negidal, speakers use Russian far more in their daily life than their heritage language, yet were obviously trying very hard to produce “correct” Negidal for the recordings which are, after all, intended

to maintain a sample of the language for posterity. There is thus considerable pressure to avoid Russian words, as shown by the relatively frequent insertion of the filler to repair Russian lexemes (nearly 37% of the repair contexts are of this type, cf. (53)). Yet speakers are not always successful in avoiding Russian lexemes: in nearly 17% of occurrences of a placeholder with an overt target, that target is a Russian word (54) or an explicit search for the Negidal equivalent of a Russian term (55). The attempt to avoid Russian lexemes is particularly striking in (30) above: after resorting to a headless relative clause to substitute for the target ‘sack’, the speaker finally gives up and uses the Russian word.

- (53) *a:ji-dgi-ja-n gun(-ə-n) e:jan e: babuf... (80) uyun-mi*
 call-REP-NFUT-3SG say-NFUT-3SG FS[why] ehm granny.R FLR-PRFL.SG
 (140) *əβəka-j jadga-s*
 granny.EVK-PRFL.SG swear.at[NFUT]-2SG
 ‘He called him back and says: “Why did you insult gran... **your whatchamacallit, your granny?**”’ [APN_babulja: 41; ID 281]
- (54) *nonon nonon bi sinə-βə sinə omugdə-βə-s=kəna (790)*
 previously previously 1SG 2SG.OBL-ACC FS belly-ACC-2SG=PTL
uyuni-kta (800) meri-kta gun-ə-n
 FLR-HORT.SG to.measure.R-HORT.SG say-NFUT-3SG
 ‘But first let me **whatchamacallit, measure** your belly.’
 [TIN_old_wolf: 46; ID 859]
- (55) *o:-ja-n tudgən (1200) uyun-duki-n (2400) təlban (480)*
 make-NFUT-3SG fast.Y FLR-ABL-3SG birch
uyun-duki-n (2700) - beresta omŋo-tca bi-si-m
 FLR-ABL-3SG birchbark.R forget-PST.PTCP AUX-NFUT-1SG
 ‘He quickly made (it) from **whatchamacallit, from** birch
whatchamacallit. - Birchbark, I forgot (the word in Negidal).’
 [AET_burunduk: 45-46; ID 28-29]

The filler is also frequently inserted when speakers correct their choice of Negidal item to a semantically more appropriate word (56) – a type of repair labelled “appropriateness-repair” by Levelt (1983: 52). Not infrequently in these cases the placeholder simply replaces the erroneous word without a more appropriate target being found (57).

- (56) *əsi=gda osoki-j oje-duki-n atali-l-tca*
 now=CONTR stove.EVK-PRFL.SG top-ABL-3SG take.off-INCH-PST[3SG]
 (1520) **uyun** (1020) *jaβa-l-tca:* *poleno-βa*
 FLR take-INCH-PST[3SG] log.R

‘Now she took off from the top of the stove... uhm... took a log (from the top of the stove).’ [TIN_boy: 5; ID 847]

- (57) *a na:nman su:n-ji sam* (0) **uyun-tca:ki-tin** (0) *kožanyj su:n-ji*
 and.R 3SG.ACC coat-INS close FLR-REM.PST-3PL leather.R coat-INS
 ‘And they had close... done whatchamacallit (covered) her with a coat, a leather coat.’ [APN_fire_water: 60; ID 383]

In (56), the speaker first produces the word *atal-*, which means ‘take off, remove’ and is used for taking off clothes, stripping bark off trees, and removing hide from a carcass; after hesitating she replaces it with *jaβa-* ‘take, grab’, which is more appropriate for the action of taking a piece of firewood from a pile. In (57) the initially chosen verb, *sam-*, means to close a door, while the required verb is *das-* ‘to cover’.²⁰ The filler also occurs in instances of ‘factual’ repair, such as in example (49) above, where the speaker repairs the number of horses from two to three.

5.4 Fillers in interaction

It has been observed in the literature (Hayashi & Yoon 2006: 499; cf. LeSourd 2003: 146–147, Keevallik 2010: 156–158) that fillers are sometimes interpreted by interlocutors as requests for help, leading to a “collaborative achievement of word search” (Hayashi & Yoon 2006: 499). While this kind of interaction is occasionally observed in the Negidal data ((58), see also (48) and (52) above), it is not very frequent: only 15 out of 364 occurrences of the filler (i.e. about 4%) in recordings in which a Negidal interlocutor was present triggered a prompt by the interlocutor.²¹

²⁰The initial choice of verb in (56) is probably due to the fact that in Russian *snimat'* can mean both ‘take off’ clothes and similar items as well as ‘take down’ something from above. In (57), the initial choice of verb might have also been influenced by Russian, since the root *kryt'* occurs both in *zakryt'* ‘close (a door)’ and in *ukryt'/pokryt'* ‘cover (with a jacket, a blanket)’.

²¹For some texts we lack the sound files of the recordings; since I was unable to verify if there was “collaboration”, these were excluded from the count, even though some of them involved more than one Negidal speaker. Given the lack of metadata for the recordings prior to 2017, the presence of an indigenous interlocutor was assessed by whether any utterances were annotated in ELAN. This would of course miss the presence of interlocutors who didn’t say anything – so that the proportion of “collaborative achievements” cited here is an upper bound.

- (58) DIN: *nu i kuyaikan ho... ho...* (1100) ***uŋun-ə-n*** (2800)
 PTL.R and.R young.human FS FS FLR-NFUT-3SG
bogdi-ji-j { GIK: *həki-sin-ə-n* } *həki-sin-ə-n*
 leg-INS-PRFL.SG step.on-TAM1-NFUT-3SG step.on-TAM1-NFUT-3SG
taj taj selafun
 DIST DIST skewer
 ‘DIN: And a child **did whatchamacallit** with his foot {GIK: stepped}
stepped on that skewer.’ [DIN_eta_duck: 11-12; ID 586]

This “collaboration” is not always successful, as shown by the following example (59). Here, after APN utters the placeholder *uŋunaβəj* ‘something for themselves’, her daughter prompts her with the term for ‘footwear’, which APN does not take up, using a term that none of the still living speakers knew, but that might refer to a specialized type of footwear.

- (59) APN: *ili taj (0) uŋun-ə-βəj (0) pa:ti-ja*
 or.R DIST FLR-DEST-PRFL.PL beat.fish.skin.ULCH-NFUT[3PL]
 { DIN: *onta-ja-βaj* } APN: *əj amusi-ja-βaj*
 footwear-DEST-PRFL.PL PROX ?-DEST-PRFL.SG
 ‘APN: Or they beat the skin for whatchamacallit for themselves. {DIN:
 footwear for themselves} For this XXX for themselves.’
 [APN_DIN_fish_skin: 99-101; ID 335]

In the context of interactional uses of the filler one might wonder whether recycling or its absence (cf. §4.1.3) is conditioned by the length of the pauses surrounding the filler: do speakers feel the need to “help” their interlocutor to follow the thread by repeating part or all of the material that accompanied the placeholder when they take a long time to retrieve the target? As can be seen from Table 4, this doesn’t seem to be the case: the minimal pause length is zero, i.e. absence of a pause, both before and after the filler, irrespective of whether there is recycling or not, and the maximal pause length is even longer after the filler when there is no recycling. The average and median pause length after a filler involved in recycling are a bit longer than when there is no recycling, but not substantially so.

To summarize, the integration of placeholders demonstrates to what extent the structure of sentences is planned before they are uttered. Furthermore, even though the Negidal filler is used primarily in situations of disfluency, the data analysed here show that it is not merely the search for rare words or for Negidal alternatives of more accessible Russian words that triggers use of the filler, since

Table 4: Comparison of pause length (in milliseconds rounded to closest tens) with and without recycling

	No recycling		Recycling	
	Pause before	Pause after	Pause before	Pause after
Minimum	0	0	0	0
Maximum	2840	8120	2820	3660
Average	690	890	730	1070
Median	500	660	640	770

ujun is frequently used when quite common or recently uttered words elude the speaker. The filler is also not a signal to the hearer that the speaker has difficulties in continuing her sentence, as shown by the rarity of “collaborative achievement” of word searches. This is congruent with the finding that filled pauses in English do not stand out from the speech stream and hence “...are not really compatible with the speaker intending to use the filled pause as a signal” (Lickley 2015: 459). Rather, there appear to be a number of factors at play that trigger filler use, such as an ongoing search for a different lexeme or the realization that the speaker has made a mistake. Unfortunately, not all of the possible triggering factors can be elucidated with the Negidal database, since most of the recordings lack videos and detailed metadata that would let us reconstruct the external context of the speech situation.

6 Discussion and conclusions

The current study has shown that the Negidal filler *ujun* is used very frequently in discourse, albeit with considerable variation across speakers. Although it can function both as a hesitative and a placeholder, the placeholder strategy predominates. Its primary function appears to be to gain time or to maintain the syntactic frame of the unfolding utterance in situations of disfluency, with pragmatic and interactive uses (avoidance or conspirational uses, turn-taking, and “collaborations” in word search) being absent or rare. Having only one item that can fulfill both strategies is cross-linguistically fairly common, being found in such varied languages as Negidal’s sister Evenki (Klyachko 2022), Russian (Podlesskaya 2010), the Caucasian languages Udi and Agul (Ganenkov et al. 2010), Amazonian Spanish (Vallejos Yopán 2023), and Tagalog (Nagaya 2022). In contrast, having

only one item that can substitute for both nouns and verbs is less common, either because languages have a placeholder that is used only for noun substitution, such as in Amazonian Spanish (Vallejos Yopán 2023) or Udi and Agul (Ganenkov et al. 2010), or because separate forms are used to stand in for nouns and verbs, as in Nahavaq (Dimock 2010) and Maliseet-Passamaquoddy (LeSourd 2003).

The placeholder strategy of the filler provides clear evidence that speakers plan the structure of their utterance in advance, before they retrieve the lexical items, since the placeholders carry the necessary inflectional morphology that integrates them into the frame of the sentence. That this planning is quite sophisticated is shown by the infrequent occurrence of mismatches between morphology on the placeholder and the target (see Figure 10 in §4.1.3). As discussed above (§5.2), these mismatches often concern “compatible” categories, such as cases that can be used with the same function and that thus “fit into” the planned frame. These are thus formally different, but semantically and syntactically congruent. The evidence from the Negidal placeholder thus strengthens the evidence from interaction errors that has been adduced in favour of structural planning in speech production (Shattuck-Hufnagel 2015: 419–421).

Several research questions have been left open and need to be addressed in the future. First of all, the current study focused mainly on the morphosyntactic aspects of the filler, with prosodic aspects being neglected; only pitch, waveform intensity, and pause length surrounding the filler were investigated (§3). Secondary markers of hesitation, such as lengthened final syllables of the filler itself and of preceding items were not taken into account, yet should be included in a full study of the phenomenon (cf. Lickley 2015: 458, and references therein: “Where “fluent” silent pause is sanctioned by prosodic structure, it is likely to be preceded by a prolonged syllable [...], and the same may apply to a hesitant pause [...]: prolongation and silence are part of the same phenomenon”).

Secondly, given that the hesitative strategy of *ujun* is quite infrequent, with less than 25% of the occurrences being categorized as such (§4.1.1), an in-depth study of Negidal disfluency would be of interest: if speakers do not use *ujun* when they experience difficulties with retrieving words or with planning their further discourse, how do they hesitate? Very preliminary data based on less than 1% of the corpus (comprising less than ten minutes of recordings from four fluent speakers) shows that the filler, including both its placeholder and its hesitative use, is about twice as frequent as other overt disfluencies, namely the hesitator *əə/ɛɛ* and the lengthening of final syllables. However, silent pauses account for ~70% of all disfluencies. These preliminary data indicate that overt, “filled” hesitation is quite rare in Negidal, but they obviously need to be confirmed with a more comprehensive study.

Furthermore, the interaction of fillers with gestures (cf. Navarretta 2016) would be worth exploring further. The individual tokens adduced in §3 and §4.1.5 show interesting differences in the timing of the gesture: the pointing gesture precedes the placeholder (example (7), Figure 4), while the gaze aversion follows after the hesitative (example (5), Figure 2), and the rolling gesture accompanies the exophoric use of *ujun* in example (36) (Figure 12). It is an open question whether these distinctions are regular, and also what different kinds of gestures accompany the different uses of the filler.

Finally, a comparison with Negidal's sister languages might provide further interesting insights into the morphosyntactic aspects of fillers, since these languages show comparable levels of morphological complexity but use fillers of different origin. Thus, Evenki (Klyachko 2022) and the Bystraja dialect of Even (Matić 2008) have dedicated fillers (*ani* and *uŋ*, respectively), while the Lamunkhin dialect of Even makes use of interrogative proforms: *iak/ia-* 'what/do what'. Preliminary data indicate that these different types of fillers might differ in how they are used, the details of which might shed more light on the diachrony and functions of fillers.

Sources

All the examples are taken from Pakendorf & Aralova (2017) and can be found by searching for the text ID. The numbers of annotation units are those that are found in the pdf-file of the transcribed, translated and glossed text. In ELAN files with multiple speakers it is best to choose the tier 'phrase-segnum-en' for the relevant speaker in the 'Grid' view, which provides the annotation number that corresponds to that given here.

Acknowledgements

I thank two anonymous reviewers for their constructive comments on a first version of this chapter, and in particular Françoise Rose, whose continued discussion and constructive criticism helped me greatly improve the paper. I also thank my colleagues Jennifer Krzonowski and Rémi Anselme for assistance: Jennifer for writing the Praat script that let me automatize most of the investigation of the intonation patterns, and Rémi for providing me with the video stills shown in some figures, for help with producing images in Praat and Excel, and for ensuring that the L^AT_EX version of this chapter is free of formatting errors. The data on which this study is based were collected with the generous support of ELDP

(grant MDP0346), which is gratefully acknowledged. I am also grateful to the ASLAN project (ANR-10-LABX-0081) of the Université de Lyon, for its financial support within the French program "Investments for the Future" operated by the National Research Agency (ANR). I furthermore thank Natalia Aralova for helpful discussion of the Negidal data not only while we were compiling the corpus, but also with respect to several examples of the filler. Last, but definitely not least, I thank all the Negidal speakers whom I had the pleasure to meet in person for their contributions to the corpus, and most especially DIN and GIK for their many hours of patient work on transcriptions, translations, and clarification of tricky questions.

Abbreviations used in glosses

1, 2, 3 indicate person, SG, PL indicate singular and plural number, respectively.

ABL	ablative	EX	exclusive
ACC	accusative	FLR	filler
ACC.INDF	indefinite accusative	FOC	focus
ADD	additive	FS	false start
ADVB.ALL	adverbial allative	HAB	habitual aspect
ADVR	adverbializer	HAB.PTCP	habitual participle
ALL	allative	HORT	hortative (1st person
AM	associated motion		imperative)
AUX	auxiliary	IDEO	ideophone
CAUS	causative	IMP	immediate future
COND	conditional		imperative (2nd
CONTR	contrastive		person)
COP	copula	IMPERS	impersonal
DAT.ESS	dative-essive	INCH	inchoative
DEC	decessive (refers to deceased people)	INS	instrumental
DEONT	deontic future	INTERJ	interjection
DEST	destinative	INTR	intransitive
DIM	diminutive	INTS	intensive
DIST	distal demonstrative	INTS.ADJ	intensive
DP	discourse particle	ITER	adjectivizer
DUR	durative	LOC	iterative
EUPH	euphemism	MULT	locative
EVK	Evenki borrowing	NEG	multiplicative
			negative auxiliary

NEG.CVB	negative converb	REP	refactive (repeated action)
NEG.FUT	future tense suffix restricted to negative auxiliary	RES	resultative
NFUT	non-future	RESTR	restrictive
NNC	borrowing from a Nanaic language	SIDE	derivative suffix with a meaning of 'side'
NUM(DAYS)	numeral derivation expressing number of days	SML	similative
OBL	oblique	SMLF	semelfactive
PASS.PTCP	passive participle	SS.ANT	same-subject anterior converb
PL.HUM	plural of nouns designating humans	SS.COND	same-subject conditional converb
PN	personal name	SS.DUR	same-subject durative converb
POSS	marker of non-canonical possessive construction	SS.SIM	same-subject simultaneous converb
PRFL	reflexive possessive	SUDDEN	sudden action
PROL	prolative	TAM	tense-aspect-mood marker
PROX	proximal demonstrative	TR	transitive
PRS.PTCP	present participle	ULCH	Ulch borrowing
PST	past tense	VAL	valency-changing suffix
PST.PTCP	past participle	VOC	vocative
PTL	particle	VR	verbalizer
Q.FUT	interrogative future	VS.PURP	variable-subject purposive converb
R	Russian borrowing	VS.SIM	variable-subject simultaneous converb
REM.PST	remote past	Y	Sakha (Yakut) borrowing
REM.PTCP	remote participle		

References

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

- Allwood, Jens, Elisabeth Ahlsén, Johan Lund & Johanna Sundqvist. 2005. Multi-modality in own communication management. In *Proceedings from the second Nordic conference on multimodal communication*, 1–20. Göteborg: Göteborg University.
- Aralova, Natalia & Brigitte Pakendorf. 2022. The causal-noncausal alternation in the Northern Tungusic languages of Russia. In Andreas Höztl & Thomas E. Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 21–62. Berlin: Language Science Press.
- Billings, Blaine & Bradley McDonnell. 2025. Fillers in the gray: Indeterminacy in their use in everyday conversation. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 211–250. Berlin: Language Science Press. DOI: ??.
- Braun, Angelika, Nathalie Elsässer & Lea Willems. 2023. Disfluencies revisited - are they speaker-specific? *Languages* 8(3). 155. DOI: 10.3390/languages8030155.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Dimock, Laura. 2010. Fillers and placeholders in Nahavaq. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 119–138. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.06dim.
- Döhler, Christian. 2025. *That* placeholder in Komnzo. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 275–313. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697589.
- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Hennecke, Inga & Wiltrud Mihatsch. 2022. Between placeholder and filler: Degrees of prosodic prominence of French *machin* and *truc*. *Pragmatics & Cognition* 29(2). 297–323. DOI: 10.1075/pc.22003.hen.
- Höztl, Andreas. 2015. A typology of negation in Tungusic. *Studies in Language* 39(1). 118–159. DOI: 10.1075/sl.39.1.05hoe.

- Jehoul, Annelies, Geert Brône & Kurt Feyaerts. 2016. Gaze patterns and fillers. Empirical data on the difference between Dutch ‘euh’ and ‘euhm’. In *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*. Copenhagen: LIU Electronic Press.
- Kalinina, Elena. 2008. Ètjud o garmonii glasnyx v negidal’skom jazyke, ili neglasnye prezumpcii o glasnyx zvukax [A study of vowel harmony in the Negidal language or unstated assumptions about vowels]. In Olga Fedorova & Ekaterina A. Ljutikova (eds.), *Fonetika i nefonetika. K 70-let’ju S. V. Kodzasova [Phonetics and non-phonetics. On the occasion of the 70th birthday of S.V. Kodzasov]*, 272–282. Moscow: Jazyki slavjanskix kul’tur.
- Kalinina, Elena. 2013. *Documentation of endangered Tungusic languages of Khabarovskij Kraj*. <http://hdl.handle.net/2196/00-0000-0000-0002-2FAD-A>.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26. 22–63. DOI: 10.1016/0001-6918(67)90005-4.
- Khasanova, Marina M. & Aleksandr M. Pevnov. 2003. *Mify i skazki negidal’cev [Myths and tales of the Negidals]*. (Endangered Languages of the Pacific Rim A2-024). Kyoto: Endangered Languages of the Pacific Rim.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hödl & Thomas Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- Klyachko, Elena. 2025. Placeholders versus general extenders in Tungusic languages. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 41–74. Berlin: Language Science Press. DOI: ??.
- Kosmala, Loulou. 2022. Exploring the status of filled pauses as pragmatic markers: The role of gaze and gesture. *Pragmatics & Cognition* 29(2). 272–296. DOI: 10.1075/pc.21020.kos.
- Kosmala, Loulou & Ludivine Crible. 2022. The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1). 216–239. DOI: 10.1177/00238309211010862.
- LeSourd, Philip S. 2003. The noun substitute in Maliseet-Passamaquoddy. In Blair A. Rudes & David C. Costa (eds.), *Essays in Algonquian and Siouan linguistics in memory of Frank T. Siebert, Jr*, 141–163. Winnipeg: University of Manitoba Press.

- Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14(1). 41–104. DOI: 10.1016/0010-0277(83)90026-4.
- Lickley, Robin J. 2015. Fluency. In Melissa A. Redford (ed.), *The handbook of speech production*, 445–474. Malden: Wiley. DOI: 10.1002/9781118584156.ch20.
- Matić, Dejan. 2008. A note on hesitative forms in two Éven dialects. In O. A. Osipova, A. V. Didenko & Elena A. Krjukova (eds.), *Sravnitel'no-istoričeskoe i tipologičeskoe izučenie jazykov i kul'tur. Materialy meždunarodnoj naučnoj konferencii XXV Dul'zonovskie čtenija [Comparative-historical and typological studies of language and culture. Materials from the international scientific conference 25th Dul'zon Readings]*, 234–239. Tomsk: Tomskij gosudarstvennyj pedagogičeskij institut.
- Matić, Dejan. 2017. Periphrastic proximatives in Éven (North Tungusic). In Agnes Korn & Irina Nevskaya (eds.), *Prospective and Proximative in Turkic, Iranian and beyond* (Iran - Turan 18), 77–93. Wiesbaden: Reichert Verlag.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- Myl'nikova, Klavdiya M. & Vera I. Cincius. 1931. Materialy po issledovaniju negidal'skogo jazyka [Materials for the study of Negidal]. *Tunguskij sbornik [Tungusic volume]* 1. 107–218.
- Nagaya, Naonori. 2022. Beyond questions: Non-interrogative uses of *ano* ‘what’ in Tagalog. *Journal of Pragmatics* 190. 91–109. DOI: 10.1016/j.pragma.2022.01.007.
- Navarretta, Costanza. 2016. The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations. In *Proceedings from the 3rd European symposium on multimodal communication, Dublin, September 17–18*, 55–61. Dublin: LiU Electronic Press.
- Oskolskaya, Sofia. 2024. The Tungusic language family. In Edward J. Vajda (ed.), *The languages and linguistics of Northern Asia. Volume 1: Language families* (World of Linguistics 10.1), 123–165. Berlin: De Gruyter Mouton.
- Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct Northern Tungusic language of the Lower Amur*. London: SOAS, Endangered Languages Archive. <http://hdl.handle.net/2196/00-0000-0000-000F-E387-1>.
- Pakendorf, Brigitte & Natalia Aralova. 2018. The endangered state of Negidal: A field report. *Language Documentation and Conservation* 12. 53–79. <http://hdl.handle.net/10125/24760>.

- Pakendorf, Brigitte & Natalia Aralova. 2020. Even and the Northern Tungusic languages. In Martine Robbeets & Alexander Savelyev (eds.), *The Oxford guide to the Transeurasian languages*, 288–304. Oxford: Oxford University Press.
- Pevnov, Aleksandr M. & Marina M. Khasanova. 2006. Negidal’cy i negidal’skij jazyk (Obščie svedenija, zvukovoj stroj, imennye časti reči) [The Negidals and the Negidal language (General information, phonology, nominal parts of speech)]. *Acta Linguistica Petropolitana* 2(1). 447–542.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Podlesskaya, Vera I. & Andrej A. Kibrik. 2009. Rečevye sboi i zatrudnenija [Speech disfluencies]. In Vera I. Podlesskaya & Andrej A. Kibrik (eds.), *Rasskazy o snovidenijax. Korpusnoe issledovanie ustnogo russkogo diskursa* [Stories about dreams. A corpus investigation of Russian oral discourse], 177–218. Moscow: Jazyki slavjanskix kul’tur.
- Podlesskaya, Vera I. & Nikolay A. Korotaev. 2022. “Étu... Kak ejo... Perepisku Éngelsa s etim... Kak ego, djavola... S Kautskim” - A case study of a placeholder in Russian. *Russian Speech = Russkaja Rec'* 5. 64–78.
- Ponsonnet, Maïa. 2025. Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 315–340. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697591.
- Rice, Alexander. 2025. *Mashti*: A multipurpose filler in Northern Pastaza Kichwa. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 373–439. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697595.
- Schegloff, Emanuel A. 1979. The relevance of repair to syntax-for-conversation. In Talmy Givon (ed.), *Discourse and syntax*, 261–286. New York: Academic Press.
- Shattuck-Hufnagel, Stefanie. 2015. Prosodic frames in speech production. In Melissa A. Redford (ed.), *The handbook of speech production*, 419–444. Malden: John Wiley & Sons.
- Vallejos Yopán, Rosa. 2023. From demonstrative to filler: *Este* in Amazonian Spanish and beyond. *Linguistics* 61(3). 651–678.
- Ventayol-Boada, Albert. 2025. Copula or placeholder: A qualitative and quantitative analysis of *xe*- in Kolyma Yukaghir discourse. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 131–168. Berlin: Language Science Press. DOI: ??.

3 A corpus-based investigation of fillers in Negidal (Northern Tungusic)

- Visser, Eline. 2025. Form and frequency of Kalamang fillers. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 251–273. Berlin: Language Science Press. DOI: ??.
- Zhao, Yuan & Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. In *Proceedings of DiSS'05: Disfluency in spontaneous speech workshop*, 179–182. https://www.isca-speech.org/archive_open/diss_05/dis5_179.html.

Chapter 4

Copula or placeholder: A qualitative and quantitative analysis of *λe-* in Kolyma Yukaghir discourse

Albert Ventayol-Boada^{ab}

^aUniversity of California, Santa Barbara ^bCurrent affiliation: Meta AI

Current grammatical descriptions of Kolyma Yukaghir analyze the form *λe-* as one of the copula roots in the language. Additionally, a variety of *λe-*'s emerge in connected speech, which are often glossed as “discourse particle” or “that one”, the latter suggesting a possible pronominal reading. Through an analysis of Kolyma Yukaghir narratives collected in the late 1990s, I show that the non-assertion uses of *λe-* in fact represent a placeholder. I also demonstrate that the distributions of both functions are distinct, and thus are best analyzed as being associated with different morphs rather than a single form with vaguely related meanings. Their synchronic resemblance, however, is probably not an accident, and the two uses are likely diachronically related. I suggest how the development of the placeholder from the copula is plausible through the reanalysis of the nominal interpretation of the copula as a referential expression standing in for another lexical item. The analysis of *λe-* in Kolyma Yukaghir provides a new potential source of placeholders cross-linguistically and contributes to the study of discourse and grammar.

keywords: copula, placeholder, discourse and grammar, Kolyma Yukaghir[‡]

1 Introduction

Analyzing grammatical morphs with identical phonological forms and distinct functions is a common linguistic task. According to Epps (2008), at least three scenarios are possible in this context. First, the similar phonological shape might be a historical accident of two or more independent forms converging. Second,



the different functions might be vaguely related and thus constitute an example of polysemy. And third, the functions might be synchronically distinct but historically related and, thus, originate from the same source. This Chapter addresses the importance of considering discourse-level features when attempting to choose among these scenarios, as discourse explanations have been shown to account for a variety of grammatical forms (Chafe 1976, Clark & Fox Tree 2002, Davis 2017, Diver 1982, Du Bois 1987, Fox & Thompson 1990, Hopper & Thompson 1980, Mithun 2008, Ono et al. 2000, Sankoff & Brown 1976, Stern 2006, Thompson & Mulac 1991; among others). Over time, salient discourse patterns may become grammatical (Ariel 2009, Couper-Kuhlen & Thompson 2008, Du Bois 2003).

Specifically, I analyze the discourse distributions of *λe-* in Kolyma Yukaghir, a form that is glossed with a variety of labels, including “copula”, “discourse particle” and “that one.” After dividing spoken discourse into Intonation Units (IU; Chafe 1979, 1992), I analyze the semantic and pragmatic information of *λe-* to establish two main functions: copula and placeholder. I then model the discourse features of each token to determine whether a polysemy analysis is possible given their distributions in discourse. Finally, I identify discourse contexts in which both interpretations are possible and can thus shed light into the historical development of the different uses of *λe-*.

Overall, this Chapter shows, first, how dividing spoken discourse into IUs can illustrate the distinct discourse-dependent functions a morph carries out in context; second, how modeling the discourse features associated with each function can establish whether their distributions are distinct; and third, how identifying bridging contexts can explain the historical relationship between both functions. In so doing, this study contributes to the growing research on fillers and their discourse features, as well as the emergence of placeholders cross-linguistically. It also provides a detailed account of a placeholder in Eurasia, and it offers methodological tools for linguists working with other endangered language communities in the documentation and description of their languages.

This Chapter is organized as follows. §2 introduces the Kolyma Yukaghir community, provides some background information on the sociolinguistic landscape and offers a typological overview of the language. §3 provides a qualitative discourse analysis of the form *λe-*, and §4 models its distribution quantitatively with random forests. §5 offers a hypothesis for the historical development of the placeholder in Kolyma Yukaghir, and §6 draws some conclusions.

2 The language and its speakers

Kolyma Yukaghir (ISO 639-3: *yux*) is one of the two extant Yukaghirc languages spoken in the Russian Arctic, the other being Tundra Yukaghir (ISO 639-3: *ykg*). Both languages are spoken in the Kolyma River basin, in the northeastern part of the Republic of Sakha (Yakutia). Kolyma Yukaghir, also referred to as Southern Yukaghir, is spoken on the Kolyma mountain range, whereas Tundra Yukaghir, also known as Northern Yukaghir, is spoken on the Kolyma River mouth by the Arctic Ocean. Figure 1 shows the Yukaghirs' historical territory at the end of the 18th century and their present-day locations.

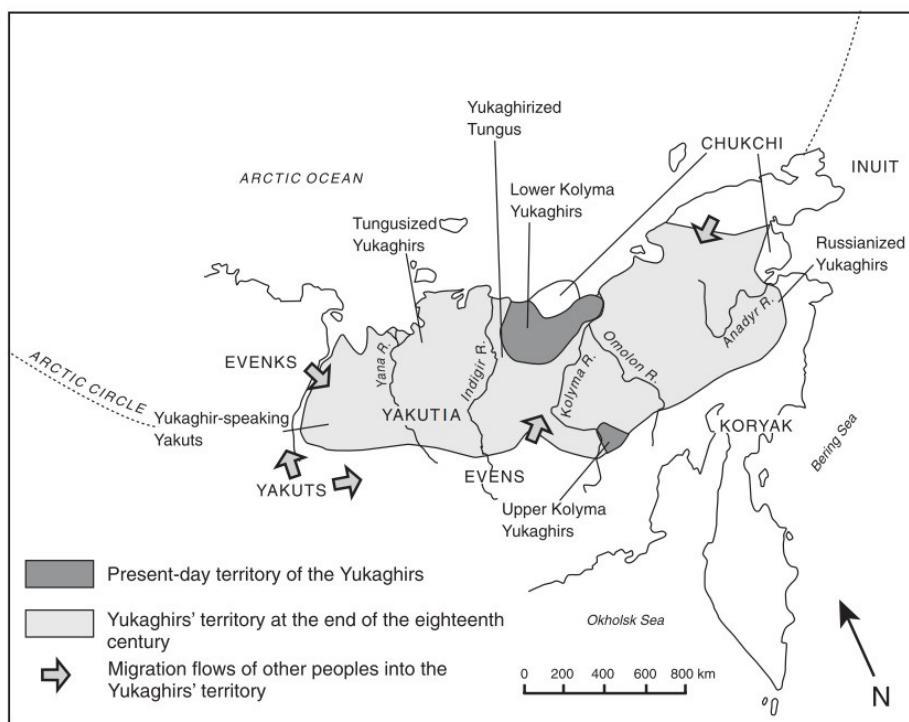


Figure 1: Map of the current and historical territory of the Yukaghirs, adapted from Willerslev (2004). The image is copyrighted by the content publisher, John Wiley and Sons, and is reproduced here under license number 5761140519254

The origin of the exonym “Yukaghir” (Russian: Юкагир) is unknown (Maslova 2003: 21), and it is the term that Yukaghirs typically use to refer to themselves when speaking in Russian, usually with a designation of their origin (i.e., Upper

Kolyma or Lower Kolyma). Alternatively, the endonyms “Odul” and “Wadul” are used as self-designations in Kolyma Yukaghir and Tundra Yukaghir, respectively. In what follows I use the term “Yukaghir” to refer to the overall ethnic group, and the term “Yukaghirc” as an encompassing denomination for both language varieties.

2.1 Language family affiliation

The Yukaghirc languages are often grouped with Paleosiberian languages (termed Paleoasiatic in the Russian literature, cf. Skorik 1986, Volodin 1997). However, there is no linguistic cohesiveness across the languages in the group (i.e., Chukotko-Kamchatkan, Nivkh, Yeniseian, and Yukaghirc). Rather, the term Paleosiberian is used to refer to the earlier presence of these groups in North Asia and the several cultural traits shared amongst them (Vajda 2009). It has been suggested that Yukaghirc might be related to the Uralic language family (Nikolaeva 1988, 2006, 2020, Fortescue 1998). This position, however, is not universally accepted (Comrie 1981, Abondolo 1998, Aikio 2014), since establishing cognates and regular sound correspondences across Yukaghirc and Uralic remains a challenging task. This potential connection to the Uralic languages is part of a wider discussion about the Yukaghirs’ origin and their relationship with their neighbors, both linguistically and culturally (Grebenuk & Fedorčenko 2018, Gogolev et al. 2022).

Currently, the two Yukaghirc languages, Kolyma Yukaghir and Tundra Yukaghir, remain as the extant members of a linguistic group that is believed to have included the now dormant Chuvan and Omok languages (Nikolaeva 2008). Kolyma Yukaghir and Tundra Yukaghir are mutually unintelligible and exhibit significant differences in their morphosyntax, phonology, and lexicon, due in part to extensive language contact with neighboring groups and multilingualism (Maslova 2003: 27–28). In fact, lexical replacements are so extensive that differentiation between both language varieties may have started as far as 2,000 years ago (Nikolaeva & Xelimskij 1997).

2.2 Sociolinguistic background of Kolyma Yukaghir

Kolyma Yukaghir is primarily spoken in the settlement of Nelemnoe, where Yukaghirs constitute a majority (Maslova 2003: 20), with a handful of speakers also living in Zyrjanka (Nagasaki 2010). The Republic of Sakha (Yakutia) grants official status to the Yukaghir languages in the areas where Yukaghirs live, alongside with the official state languages, that is, Russian and Sakha/Yakut (Grenoble

2022). The law, which also grants similar status to four other Indigenous languages in the Republic, aims to create the necessary conditions for their preservation and development (Vinokurova et al. 2022).

Despite this legal recognition, Kolyma Yukaghir is considered critically endangered and remains the language of communication only among the oldest generation (Grenoble 2003, but see Kazakevich et al. 2022 for a critique of vitality labels). According to Maslova (2003: 20), in the late 1990s Russian was typically the first language for people under the age of 60, although many declared Kolyma Yukaghir as their mother tongue. The youngest generation at the time was virtually monolingual in Russian, according to the author, although Kolyma Yukaghir had been taught in the local school in Neleminnoe since the mid 1980s. This pattern thus suggests a rapid decline in Kolyma Yukaghir language use and transmission.

However, Maslova (2003: 22–23) also points out that a similar pattern was already documented by Jochelson (1926) at the end of the 19th century. This similar attestation is striking and raises the question of the processes that have maintained language use over the span of a century, if the youngest generation were not to acquire the language, as suggested in both accounts. Maslova (2003) hypothesizes that language users might underreport their own and others' language proficiency for various reasons. Similarly, she observes that middle-aged community members who had previously claimed not to know the language in fact used it later in life.

These complex sociolinguistic patterns are closely intertwined with several structural changes under way in Kolyma Yukaghir. These changes affect predominantly the language's morphosyntax, and they reflect a clear influence from Russian grammatical structures (Matić 2008)—an influence that is also documented in neighboring languages (Grenoble 2000, Pakendorf 2024, Anderson 2017, Kantarovich 2020). Similarly, Maslova (2003: 24) notes that the language used by each generation is clearly distinct, with younger language users making wide use of code-mixing and incorporating Russian loanwords without phonological adaptation. It is important to consider these changes because they are likely to involve word search processes, a common context in which speakers might deploy placeholders.

2.3 Typological overview of Kolyma Yukaghir

Like other languages in the Siberian linguistic area (Anderson 2006, Pakendorf 2024), Kolyma Yukaghir is strongly head-final, and it displays SV/AOV constituent order with nominative-accusative alignment. Additionally, Kolyma

Yukaghir, like its sister language, is known for the grammaticalization of information structure in the languages' syntax, a property often referred to as *grammatical focus*, *focus system* or *focus construction* (Comrie 1981, Krejnović 1982, Maslova 1997, 2003, Nagasaki 2018, Nikolaeva & Xelimskij 1997). Morphologically, Kolyma Yukaghir is a predominantly agglutinating, suffix-dominant language, with partially fusional morphology. Suffixes display some allomorphy due to residual vowel harmony and consonantal assimilation processes (Krejnović 1982, Maslova 2003, Nagasaki 2010). Examples in this Chapter are given in the practical orthography first, followed by an IPA transliteration that also includes morpheme boundaries marked with hyphens.

The number and identity of parts of speech in Kolyma Yukaghir remains a contested issue (Ventayol-Boada et al. 2023). Morphologically, roots display a certain degree of flexibility and whether they are used "verbally" (i.e., for predication) or "nominally" (i.e., for referring) depends on the affixes attached to them. The number and range of affixes with which a given root occurs varies for each use, and, thus, words exhibit different levels of morphological complexity. Roots used "verbally" have the largest number of affixal slots, some of which can be filled by a wide range of possible items (e.g., the slot for aspect). For the most part, boundaries between slots are well-defined, but some of the affixes for person have fused with the preceding morphemes. Figure 2 shows the template for words used "verbally" to denote assertions;¹ only the root, assertion and person slots are obligatory. Converbs and participles follow a different morphological template.

Roots used "nominally", on the other hand, have fewer slots. Typically, slots in the nominal template can be filled by fewer possible affixes than roots used "verbally" and sometimes occur without affixes at all. In discourse, plurality is not obligatorily marked, and terms designating plural referents may not be marked with the plural suffix. There is no case agreement between nominal referents and their modifiers. Figure 3 shows the template for words used "nominally."

There are three demonstrative stems: *ti-* (proximal), *adaa-* (medial), and *taa-* (distal, typically invisible) (Maslova 2003: 238–248). These stems can bear the

¹I use the term *assertion* for speech events in which "the speaker describes or makes a claim concerning a general state of affairs" (Vatanen 2014: 3). Speech events that express questions and commands are marked differently (cf. Silverstein 1976, van Valin 2003). Maslova (2003) classifies the assertion-making morphemes according to valency (i.e., intransitive vs. transitive) and as to whether they highlight the event (i.e., verb-focus) or the event participants (i.e., subject-focus or object-focus). However, according to Ventayol-Boada (2023), this description fails to account for a significant number of utterances in discourse in which "intransitive" forms are attested with two participants, and 'transitive' forms are found with a single participant. Since their different discourse functions remain unclear, I gloss the four forms as Ass "assertion."

Root	Aspect	Evidentiality	Number	Tense	Assertion	Person
Root	-oo (RES)	- <i>əl</i> (EV)	- <i>ŋi</i> (PL)	- <i>te</i> (FUT)	- <i>je</i>	Person
	-ie/aa (INCH)				- <i>l</i>	
	-ej/aj (PFV)				- <i>me_E</i>	
	-nu (IMPF)				- <i>me_P</i>	
	-nun (HAB)				etc.	

Figure 2: Verbal template in Kolyma Yukaghir, adapted from Maslova (2001)

Root	Number	Possessor	Size	Case
Root	- <i>p/pe/pul</i> (PL)	- <i>ki/gi</i> (3POSS)	- <i>die</i> (DIM)	- <i>kele/gele</i> (ACC)
		- <i>de</i> (3POSS)	- <i>tege/tke</i> (AUG)	- <i>ge</i> (LOC) - <i>get</i> (ABL) etc.

Figure 3: Nominal template in Kolyma Yukaghir, adapted from Maslova (2003)

linker *-ŋ* when modifying a referent, and a different set of suffixes when used adverbially to denote source, direction or time. The anaphoric function is expressed by the forms *tuən*, *aduən*, and *tamun*, respectively.

2.4 Data

For this study I analyzed 34 of the 41 texts collected by Nikolaeva & Mayer (2004) in the late 1990s. The texts were stripped of glosses, transliterated into Cyrillic orthography, and divided into Intonation Units (see §3.2) using the audio recordings.² In total, the corpus contains 10,258 words with 1,099 different types. These 34 texts were narrated by seven different community members, although one speaker, Vasilij Šalugin, contributed the majority, with 22 texts.

All texts are monologic, although they differ slightly in genre. They include personal stories, folktales, legends, and what can be characterized as “sharing knowledge” narratives (Smith 1999), which include knowledge and information about the Yukaghirs’ worldview, ranging from traditions to games, to fortune-telling practices, among other topics. Many texts, however, combine elements from different genres. For example, legends often include retellings of histori-

²All the corpus texts, annotations and script files are available on GitHub at https://github.com/aventayloboada/placeholder_in_Odul.git

cal events and events in the speaker's life, whereas knowledge narratives can incorporate personal experiences without them being the focus of the story.

3 Qualitative analysis: The functions of *λe-*

3.1 *λe-* as copula

Current grammatical descriptions of Kolyma Yukaghir analyze the form *λe-* as the root for one of the two copulas in the language, the other being *oo-* (Krejnovič 1982, Maslova 2003, Nikolaeva 2006, Nagasaki 2010). Copulas are typically associated with connecting a referent to a predicational, identity or existential clause (Citko 2014). All three uses are attested in Kolyma Yukaghir: (1) shows a predicational use, in which the copula ascribes a property to a referent; (2) demonstrates an identity usage, by which the copula connects the terms for fishing rod in Kolyma Yukaghir and Russian; and (3) displays an existential meaning, in which the copula introduces the hero of the narrative. Additionally, *λe-* can function as an auxiliary following a converb, as shown in (4).

- (1) *Маџилги атахун нүгэннүэй чомоолбэн хаарэ*
mažil-gi ataq-u-n nugen-ne-j tʃomooolben qaar-e
coat-3POSS two-(EP)-LNK arm-PROP-PTCP elk skin-INS
λe-j-bed-ek.
λe-j-bed-ek
COP-PTCP-NMLZ-PRED
'His coat was (made) with two-finger (thick) elk skin'
(Nikolaeva & Mayer 2004, 36:10)
- (2) *Уðочкаðоон луусьии тиñэ λьэт.*
udočka-doon luuṣii tite λe-t
fishing.rod-NMLZ Russian like COP-CVB.CTX
'Fishing rod is udochka in Russian'
(Lit. "Being like an udochka in Russian") (Nikolaeva & Mayer 2004, 50:17)
- (3) *Taam одун ханисъэ λьэлъэл.*
taa-t odu-n qanjiče λe-λel-Ø
there-ADV.ABL Yukaghir-LNK hunter COP-EV-ASS.INTR.EF.3SG
'There was a Yukaghir hunter' (Nikolaeva & Mayer 2004, 21:1)

- (4) *Йүгүдин лъэйэ.*
 jug-u-din ле-je
 kiss-(EP)-CVB.PURP COP-ASS.INTR.EF.1SG
 'I wanted to kiss (her)' (Lit. 'I was to kiss')
 (Nikolaeva & Mayer 2004, 40:70)

The copula typically occurs in clause-final position, as expected in a strongly head-final language. Additionally, the referent connected to the predicational, identity or existential clause typically bears no overt morphology, except when *λe-* is marked with one of the participant-focus assertion forms. In these cases, the referent is typically marked with the predicative case marker, as shown in (5).

- (5) *Mum aархaa иркин шилупкак лъэл.*
 mit aarqaa irk-i-n fl'upka-k le-l
 1PL at one-(EP)-LNK boat-PRED COP-ASS.INTR.PF.3SG
 'There was a boat by us' (Nikolaeva & Mayer 2004, 50:77)

However, the analysis of *λe-* as a copula is striking given some of its uses attested in discourse. For example, in (6) the first token of *λe-* is marked with an assertion but it co-occurs with a case marking form on the neighboring pronoun *tamun*, which refers to the personification of the devil into a girl, not typically associated with copulas (i.e., accusative case). Similarly, the second *λe-* token does not signal (and it is not marked for) an assertion at all and takes case morphology, which is typically associated with the nominal class. However, Nikolaeva & Mayer (2004) gloss both uses as a copula.

- (6) *Тамунгэлэ лъэлъэлум түдэ лъэгэлэ*
 tamun-gele λe-ləl-u-m tude λe-gele
 that-ACC COP?-EV-(EP)-ASS.TR.EF.3SG 3SG.GEN COP?-ACC
истриэлэгэлэ миндэллэ табудэ дьэ аийиипъэлум.
 istriela-gele min-delle tabud-e dze ajii-λel-u-m
 arrow-ACC take-CVB.SEQ that-INS DP shoot-EV-(EP)-ASS.3SG
 'He took an arrow and shot her' (Nikolaeva & Mayer 2004, 23:5)

Additionally, an identical form *λe-* is glossed as "that one" by Nagasaki (2010), as shown in (7) from the folktale "Why the crow is black." This analysis suggests the existence of a pronominal *λe-*, but she does not include the form in the list of pronouns in the language. Like the second *λe-* in (6), the token below is also marked with case (i.e., dative).

- (7) Пугэл, пугэл, пугэл, **льЭүин** миэстэүин
 puge-l puge-l puge-l ле-ŋin mieste-ŋin
 be.hot-PTCP be.hot-PTCP be.hot-PTCP **that.one-DAT** place-DAT
 мэрэйди.
 mer-ej-ŋi-∅
 fly-PFV-3PL-ASS.INTR.EF
 'They flew away to the hot, hot, hot place' (Nagasaki 2010: 252)

Finally, there appears to be a discourse particle that also takes the shape *λe*, based on Nikolaeva & Mayer (2004)'s glossing. Example (8) from the story "Tobacco", in which the speaker narrates his experience with this substance from an early age, illustrates a case of such usage. Unlike previous examples, *λe* in (8) bears no additional morphology.

The label ‘discourse particle’, however, is not particularly insightful as to the functions *λe* carries out in (8). Unraveling its functions is hampered by Nikolaeva & Mayer (2004)’s additional list of four discourse particle forms that start with *λe*. These are: *λege*, *λegen*, *λedegen*, and *λelek*. It is plausible to assume that these four forms are morphologically complex. What remains after singling out *λe-* in these four forms coincides with affixes found in the nominal paradigm: -*ge*, -*gen* and -*lek* are the locative, prolative and predicative cases, respectively; whereas -*de-* marks 3rd person possession. Thus, all these ‘discourse particles’ can be seen as involving the form of *λe-* followed, in some cases, by case marking, a description that is strikingly similar to what is attested in examples (6-7) above despite the different glosses offered: copula in (6) and ‘that one’ in (7). These structural similarities suggest that these additional uses of *λe-* are related but non-copular. The question as to what their functions might be, however, remains unanswered.

3.2 *Ae-* as placeholder

One possibility to investigate the non-copular uses of *λe*- is to turn to discourse and analyze its semantic and pragmatic components through extended stretches of naturalistic discourse. Discourse-level explanations have been shown to account for a variety of grammatical forms, including relative clauses (Sankoff

& Brown 1976, Fox & Thompson 1990), transitivity (Hopper & Thompson 1980, Thompson & Hopper 2001), case and grammatical relations (Diver 1982, Ono et al. 2000), ergativity (Du Bois 1987), epistemic parentheticals (Thompson & Mulac 1991), hesitators (Clark & Fox Tree 2002), reflexives (Stern 2006), dependency markers (Mithun 2008), adverbial clitics (Davis 2017), to name but a few.

To do that, a useful first step is to divide texts into Intonation Units (IU; Chafe 1979, 1992), defined as “a stretch of speech uttered under a single coherent intonation contour” (Du Bois et al. 1993: 47), or as Chafe (1994: 29) puts it, the “spurts of language” in which speakers typically produce speech. The boundaries of IUs are defined in terms of prosodic features and, although they tend to overlap with the boundaries of syntactic units (Du Bois 2003, Himmelmann 2022), IU boundaries are established by identifying a variety of phonetic cues. These include: pauses, pitch resetting, lengthening, and tempo shifts on the audio recordings (Chafe 1992, Du Bois et al. 1993). These criteria make IUs identifiable cross-linguistically (Himmelmann et al. 2018, Troiani 2023).

The role of *λe-* indeed becomes more apparent when dividing Nikolaeva & Mayer (2004)'s texts into IUs. Consider example (6) above, reprinted below in IUs as (9). Two features become apparent. First, the form *λe-* appears in both cases at the end of an IU, and, in both cases, it is followed by a substantial pause: 0.4 and 0.5 seconds, respectively. The second important feature is that the morphology attested on *λe-* is also found on two lexical items that appear later in the discourse. These are: *istriela* ‘arrow’ and *ajii* ‘to shoot’. Put together, these two features suggest that *λe-* in (9)³ is working as a placeholder. The same analysis can be made for (7-8), in which *λe-* is also found with the same morphology as lexical items that appear immediately after in discourse: the dative case in *mieste* ‘place’ and assertion zero-morpheme in *ooze* ‘to drink’.

- (9) *Тамунгэлэ λьэлъэлум,*
tamun-gele λe-fel-u-m
that-ACC PH-EV-(EP)-ASS.TR.EF.3SG
‘He watchamacallited that’
(0.4)

³In the examples, each new line represents a new IU, and pauses are given in seconds within parentheses. Additionally, I use the following symbols from Du Bois et al. (1993):

- . Final intonation
- , Continuing intonation
- Truncated/abandoned IU

түдэ **лъэгэлэ**,

tude **λe-gele**

3SG.GEN PH-ACC

‘His watchamacallit’

(0.5)

истриэлагэлэ миндэллэ табудэ ðвэ,

istriela-gele min-delle tabud-e ðze

arrow-ACC take-CVB.SEQ that-INS DP

‘After taking (his) arrow, with that’

айшилъелум.

ajii-λel-u-m

shoot-EV-(EP)-ASS.TR.EF.3SG

‘He shot (her)’

(Nikolaeva & Mayer 2004, 23: IU 26–31)

Hayashi & Yoon (2010: 37) define placeholder in the following terms: “(i) it is a referential expression that is used as a substitute for a specific lexical item that has momentarily eluded the speaker (and which is often specified subsequently as a result of a word search), and (...) (ii) it occupies a syntactic slot that would have been occupied by the target word, and thus constitutes a part of the syntactic structure under construction.” As mentioned, the uses of *λe-* in (9) fulfill both properties of this definition, since both tokens anticipate a word that is retrieved later in discourse (i.e., ‘arrow’ and ‘shoot’), and both are integrated in their syntactic slot—so much so that they are already marked with the relevant morphology of the lexical item they stand in for.

Since roots in Kolyma Yukaghir display a certain degree of polycategoriality and can be used “verbally” or “nominally” depending on what affixes attach to them (Ventayol-Boada et al. 2023), it is not surprising that *λe-* in (9) can fill in the place of a word used for predication or for referring. The ability of a single placeholder to stand in for multiple categories (especially nouns and verbs, but sometimes also adjectives or adverbs) has been found in other languages of Eurasia (e.g., Nganasan, Ulcha and Udihe; Podlesskaya 2010), as well as Maliseet-Passamaquoddy (LeSourd 2003) and Indonesian (Wouk 2005).

More interesting is that *λe-* is attested substituting demonstratives, since demonstratives are often the source of placeholders (Hayashi & Yoon 2010, Podlesskaya 2010). Demonstrative targets are not the most common, but the texts collected by Nikolaeva & Mayer (2004) display 10 such cases. In example (10), *λe-* bears the linker *-ŋ*, which only attaches to demonstratives to grammatically mark their relationship with their referent.

- (10) **льэй,**
λε-η
PH-LNK
‘Whatchamacallit’

(1.2)

<i>тaη</i>	<i>nулумt</i> ,
<i>ta-η</i>	<i>pulut</i>
<u>that-LNK</u>	<u>old.man</u>
<u>‘That old man’</u>	
<i>омулge</i>	<i>иахайə</i> ,
<i>otul-ge</i>	<i>jaqa-je</i>
<i>camp-LOC</i>	<i>reach-ASS.INTR.EF.1SG</i>

‘I arrived at the camp (of that old man)’

(Nikolaeva & Mayer 2004, 34: IU 201–204)

In addition to the variety of categories *λe-* can stand in for, examples (9) and (10) also show that the placeholder fully mirrors the grammatical categories marked on its target. However, this need not be the case, since instances of zero or partial mirroring are also attested. Consider example (11) below, from the story “A game”, in which the speaker narrates the rules of a four-hand game between two people. In this excerpt, the narrator is engaging in a word search while, at the same time, putting together the relevant morphosyntactic pieces as they relate to the “pinkie finger.” Specifically, he is trying to recall *numu-* ‘to press’ or ‘to grab’, which has eluded him temporarily, and to mark it with the relevant morphology. A first placeholder is uttered after a 0.9-second pause, but it does not bear any morphological marking.⁴ Immediately after, a second placeholder is uttered, this time with most morphological information included. At this point, the morphosyntactic scaffolding is almost complete, but the lexical retrieval has not succeeded yet. After a second-long pause, the lexical item *numu-* is finally produced, with the addition of the causative and perfective suffixes.

⁴The lack of morphology on the first *λe-* in (11) raises the question as to whether it could be best analyzed as a hesitative. The analysis is hampered by the existence of zero-morphemes both in the nominal case system and the verbal assertion paradigm (cf. the reanalysis of example (8) at the beginning of §3.2). In total, I identified 26 examples of *λe-* targeting a lexical item that does bear a zero-morpheme, which represents an 8.5% of the total number of placeholders. Example (11) is the only one with two *λe-*'s produced sequentially.

- (11) *Aǔ myđə үукоол пиэдисъэ,*
aj tude juk-oo-l piediçe
also 3SG.GEN small-RES-PTCP finger
'Also, his pinkie finger'

(0.9)

- аү лъэ,*
aj ле
also PH
'Also, whatchamacallit'

лъэнульэлмэлэ.
λe-nu-λel-mele
PH-IMPF-EV-ASS.TR.PF.3SG

'He whatchamacallits'

(1.0)

нүумышэйнульэлмэлэ.
numu-f-ej-nu-λel-mele
press-CAUS-PFV-IMPF-EV-ASS.TR.PF.3SG

'He grabs (his pinkie finger too)' (Nikolaeva & Mayer 2004, 43: IU 40–45)

However, the search for the target lexical item is not always successful, either because the relevant item cannot be retrieved or because it simply does not exist. Thus, speakers might sometimes deem the placeholder as a sufficient reference in a given communication context. Podlesskaya (2010) calls this use of placeholders as an “approximate naming” function. This use can be seen in (12) below. This example comes from “The singing girl” story, which retells the events also described in the story “The shaman Staryj” mentioned above, albeit from the perspective of a different community member. In this version, a statue was erected in the Popovka river basin where the demonic girl was killed. However, this statue was later lost. In (12), the placeholder stands in for the statue, but the lexical item is never retrieved. These are, in fact, the very last IUs in the narrative. Rather, the speaker considers the placeholder to be enough information given that the referent has been established and is the focus of discussion.

- (12) *Шаал огбоой,*
ʃaal oɣ-oo-j
tree stand-RES-ASS.INTR.EF.3SG
'The tree stands [there]'

(0.8)

это,

eto

this

'This'

(1.0)

ЛъЭГИ *өйлөв.***λe-gi** *øjλe***РН-3POSS NEG**

'There is no whatchamacallit'

(0.8)

И всё.

i vsjo

and all

'That is the end'

(Nikolaeva & Mayer 2004, 28: IU 305–311)

Most likely the use of the Russian demonstrative *eto* ('this') in (12) is also an instance of deploying a placeholder, since that is the main placeholder used in Russian (Podlesskaya 2010). The second-long pause following *eto* suggests that a word search may have been carried out, but only the native placeholder was found. At the time of speaking, however, it was deemed sufficient. An analysis of borrowed or code-switched placeholders and their relationship with native forms is clearly called for, but is beyond the scope of this chapter. See Visser (2025 [this volume]) and Klyachko (2025 [this volume]) for more observations on borrowed placeholders in Kalamang and Tungusic languages, respectively.

Example (12) also reveals that placeholders are not always followed by an IU boundary, as (9) through (11) might suggest. In fact, the placeholder and its target can be found in the same IU, as shown in (13), where *λe-* is immediately followed by the target 'Russian house' (i.e., the Yukaghir term for the town of Vekhnekolymsk). The co-occurrence of placeholder and target in the same IU is not entirely surprising; as Chafe (1994: 63) points out, "people sometimes revise their choice of wording while an intonation unit is already in progress."

- (13) *Məm amaxun,*
 met ataq-u-n
 1SG two-(EP)-LNK
 'My two'

(0.3)

чaaачaa эрэ,
tʃaatʃaa ere
elder.brother only
'Elder brothers only'

(0.9)

льягэ луусьин нумөгэ модоуи.
λe-ge luučii-n numø-ge modo-ŋi-Ø
PH-LOC Russian-LNK house-LOC live-3PL-ASS.INTR.EF
'Lived in whatchamacallit, in Verkhnekolymsk'

(Nikolaeva & Mayer 2004, 49: IU 10–15)

To summarize, dividing spoken discourse into IUs helps to reveal that the non-copular use of *λe-* is indeed an instance of a placeholder. As such, it can stand in for a variety of lexical classes (i.e., at least nouns, verbs, and demonstratives) and typically mirrors, either partially or fully, the grammatical categories marked on the target when it is retrieved after a word search. However, speakers may deem at times that the placeholder is sufficient and do not seek further clarification of the intended target. Finally, the placeholder can be followed by an IU boundary and maybe a pause, although the target may be recovered immediately and thus be integrated in the IU in progress.

3.3 Characterizing *λe-*

At this point, a question remains: given their identical phonological form, what is the relationship between these distinct functions (i.e., copula and placeholder)? As Epps (2008) points out, this is the most recurrent issue that descriptive linguists face. According to her, there are three potential scenarios worth considering in this situation: is the resemblance a historical accident of two or more independent forms converging into the same phonological shape? Is it a case of polysemy (i.e., a single form with meanings vaguely related)? Or is it perhaps the case that, synchronically, their functions have diverged so much that there is no longer a discernible semantic relation, even though the forms are historically related?

As shown in examples (9) through (13), the two uses of *λe-* can be identified under the prototypical roles of copula and placeholder and, thus, are best described as synchronically distinct. To investigate whether a polysemy analysis is still possible in this scenario, one possibility is to analyze their distributions in discourse.

Specifically, one can ask whether each function is associated with different discourse features, under the assumption that related senses of a word display similar distributions. Discourse features, such as intonation contour, preceding and following pause lengths or number of words in an IU, can offer predictions to characterize each use of *æ-*. These predictions can be made explicit through the use of quantitative methods, which offer a probabilistic characterization of the discourse features that tend to co-occur with each of the functions of *æ-*, and, crucially, evaluate whether or not the features in each case are statistically distinct from each other. §4 provides an overview of the methods and data used in the quantitative analysis, as well as the variables and their predictions.

4 Quantitative analysis

4.1 Data

From the texts collected by Nikolaeva & Mayer (2004) that I analyzed, I extracted all 453 tokens of *æ-*. I annotated them for their function (i.e., copula vs. placeholder) through a qualitative analysis of the discourse context they appear in. The criteria to assign a token to a function were based on the prototypical role(s) associated with each function: for the copula, to connect a referent to a predicational, identity or existential clause (Citko 2014); for the placeholder, to momentarily replace a target in its syntactic slot (Hayashi & Yoon 2010). In total, the tokens include 148 copulas and 305 placeholders. The placeholder tokens show substantial variation in terms of their transparency as placeholders. Of the 305 placeholders, 219 tokens were followed by the target lexical item within the next five IUs (i.e., 71.8%). Of these, in 146 examples the morphological categories marked on the target were fully mirrored on the placeholder (i.e., 66.6%).

Overall, the copula occurs 14.43 times per 1,000 words, although with substantial differences across the seven speakers: from no copulas uttered by Anna Šadrina to 31.35 copulas per 1,000 words by Fevron'ja Šalugina. As for the placeholder, it occurs 29.73 times per 1,000 words. Here, too, are important differences across the seven speakers, ranging from 17.28 placeholders per 1,000 words by Ivan Dolganov to 46.23 by Dmitrij Djačkov. Similar interspeaker variation is attested in Besemah (McDonnell & Billings 2025 [this volume]), Dalabon (Ponsen 2025 [this volume]) Kalamang (Visser 2025 [this volume]), and Negidal (Pakendorf 2025 [this volume]).

The placeholder frequency represents a much higher average than the reported frequencies in Russian elicited narratives (as reported in Podlesskaya 2010 from Podlesskaya & Kibrik 2006) and Mandarin conversations (Zhao & Jurafsky

2005), with 5 and 6.68 placeholders per 1,000 words respectively. The overall use of placeholders in Kolyma Yukaghir discourse, however, is presumably even higher, since these counts exclude the use of Russian placeholders, which are also attested, as shown in (12), but not analyzed systematically here. It is possible that this higher use of placeholders reflects the complex sociolinguistic patterns in the Kolyma Yukaghir community. The attrition and interference from Russian underlying the strategy of “survival through modification” (Maslova 2003: 24) creates shifting linguistic repertoires across generations. In this context, words might temporarily elude speakers more often, thus explaining a higher need to deploy a placeholder when compared to a non-endangered language.

4.2 Methods

To predict the probability of *xe-* functioning as a copula or placeholder based on its discourse context, I retrieved six discourse features for each token. These features include: the position of *xe-* in the IU (hence *match position*), the number of words in the IU, the IU intonation contour, the lengths of preceding and following pauses (if any), and the speaker of the text in which the token was found. The match position was later normalized to a range between 0 and 1, where 0 indicates the match was in the first position, and 1 marks the last position. For tokens which were the only word in the IU, the match position was normalized to 1. While coding the match position in such examples as 0 is also possible, the alternative is more informative, because it captures that the match is followed by an IU boundary like other IU-final matches. The goal is to measure how each function co-varies with each feature (i.e., to assess the relationship between the two uses of *xe-* by examining how changes in the discourse features are associated with each function).

These discourse features yield different expectations for each function. First, the copula is expected to occur later in an IU than the placeholder, and thus to be associated with a higher value of match position. Kolyma Yukaghir is strongly head-final, and thus the copula, as it bears an assertion-making form, is more likely to appear towards the end of an utterance, followed by an IU boundary. The placeholder, too, is often found in an IU-final position: if a speaker is performing a lexical search, they might stall and project an IU boundary. However, the placeholder can also be followed immediately by its target in an IU in progress, as shown in (13), or be deemed as a sufficient reference, as shown in (12). Since the placeholder can stand in for lexical items performing different functions (i.e., nouns, demonstratives, verbs, etc.) and these might appear in different positions

within an IU, the overall expectation for the placeholder is to be associated with a lower value of match position.

Second, the copula is expected to appear in more fluent discourse and longer IUs than the placeholder, thus yielding a higher value for the number of words in an IU. The placeholder, on the other hand, is more likely to be involved in a word search process, and thus with disfluencies (Zhao & Jurafsky 2005) and overall shorter IUs. As a result, the expectation is for the placeholder to be associated with a lower value for the number of words in an IU.

Third, the copula is expected to mark the end of an assertion, and thus to be associated with a drop in pitch and a falling intonation contour. On the other hand, the placeholder is expected to mark a word search problem, and therefore it is more likely to co-occur either with a less pronounced drop in intonation or with the speaker abandoning the IU altogether. The expectation for the placeholder is thus to be associated with continuing and truncated intonations.

Fourth, the copula is expected to co-occur with shorter delays than the placeholder since no repair is involved, thus yielding a lower value for preceding and following pauses. Longer following pauses are also possible, however, and may result from speech planning, as the speaker finishes the current chunk of discourse and moves on to the next item they want to communicate. The placeholder, on the other hand, is more likely to co-occur with longer delays before and after it is deployed, as word search and repair processes correlate with unfilled pause lengths (Clark & Fox Tree 2002). Thus, the expectation is for the placeholder to be associated with a higher value for preceding and following pauses.

Finally, individuals are expected to vary in terms of how likely they are to use *æ-* as a copula or a placeholder. As mentioned in §4.1, the number of copulas and placeholders varies substantially from person to person, which might reflect individual idiosyncratic preferences (Clark & Fox Tree 2002). For example, both Anna Šadrina and Fevron'ja Šalugina contributed one text each, but Anna Šadrina's seven *æ-*'s are all instances of the placeholder, whereas Fevron'ja Šalugina's twenty *æ-*'s are split evenly with ten tokens for each function.

For the modeling, I used a classification random forest (Breiman 2001) with the function of *æ-* (i.e., copula vs. placeholder) as the dependent variable, and the discourse features outlined above as well as the speaker⁵ as the independent

⁵At this point, random forests do not have random effects capability, which are typically used to control for speaker effects. Treating the speaker information as a fixed effect can account for the attested variation in the data, but limits the model's generalizability to new speakers. In the context of language endangerment, such generalizability may be less concerning, as texts from new people are less likely to be included in a new analysis.

variables. Random forests are a machine learning method that makes predictions by combining the results of many decision trees. Each tree resembles a flowchart, where each decision point evaluates different attributes of the data. To build each tree, the data is split into two different parts: one trains the model, and the other tests it. The training data for each tree is sampled randomly with replacement (aka *bootstrapping*), that is, some data points may appear multiple times while others may not appear at all. During training, the algorithm also randomly omits some predictors to determine which ones contribute the most. In the end, the predictions from all the trees are averaged together to produce the final result, and the prediction accuracy is calculated by comparing these predictions to the actual observations (i.e., the attested label of the dependent variable for each data point, in this case copula or placeholder).

Random forests are particularly suitable for the analysis because they can handle data structures with non-Gaussian (i.e., non-normal) distributions, and they are applicable to small- n large- p (i.e., few data points with many predictors) scenarios while avoiding problems of collinearity (i.e., the scenario in which two or more predictors are strongly related) (Gries 2021). However, detecting and interpreting interactions in random forests requires caution. The importance of a predictor as shown by the random forest does not necessarily reflect an individual effect, since its importance can be the result of a predictor's participation in an interaction (i.e., the effect of two or more variables that is greater than or different from what would be expected from adding up their individual effects) (Gries 2021).

One solution to this problem is to include the combined effect of two or more predictors that one is interested in into the random forest modelling, which Gries (2021) recommends following Forina et al. (2009), Strobl et al. (2009) and Baayen (2011). In this case, a potential interaction of interest is between the match position and the number of words in an IU: the copula's higher likelihood to appear towards the end of an IU and its occurrence in an IU with more elements are dependent on discourse fluency and the lack of a word search process. Thus, I created an interaction predictor by multiplying the normalized match position with the number of words in the IU and included it in the model.⁶ Tables 1 and 2 summarize the seven independent variables used to annotate each token of *λe-*.

⁶The multiplicative interaction becomes 0 when the normalized match position is 0 (i.e., when the match is in the first position in an IU) regardless of IU length, thus suggesting that information about IU length is lost in such cases. However, modifying the first position in an IU to 0.1 in the normalization process to avoid multiplying by 0 and preserving IU length information in the interaction does not change the modeling results.

Table 1: Categorical variables and their levels

Categorical predictor	Levels
Intonation contour	Continuing, Falling, Truncated
Speaker	Vasilij Šalugin, Anna Šadrina, Ivan Dolganov, Fevron'ja Šalugina...

Table 2: Numeric variables and their summary statistics

Numeric predictor	Min	Max	Mean	Median
Norm'd match position	0	1.0	0.83	1.0
Words in IU	1	8.0	2.08	2.0
Interaction: Position \times Words	0	5.0	1.60	1.0
Preceding pause (in sec)	0	7.9	0.81	0.5
Following pause (in sec)	0	5.0	0.80	0.5

Despite the long tail in the pause length distributions, their logs do not alter the results, since the log transformation is monotonic (i.e., it does not fundamentally alter the relationship between data points). Therefore, the pause length measurements are left in seconds for simplicity.

4.3 Results

Due to the unbalanced distribution of the two forms, the baseline/no-information rate accuracy of the classification model is already at 67.33%. That is, without any information the model would predict the placeholder 67.33% of the time (i.e., its overall distribution with respect to the copula). The model performs with a 74.17% true prediction/out-of-bag accuracy, and with a 74.08% as the out-of-bag Area Under Curve (AUC, the equivalent of the C -score in regression modelling). The difference between the baseline and the prediction accuracy is statistically significant ($p_{binomial\ test} < 0.001$). Figure 4 shows the Mean Decrease Accuracy plot, which indicates the importance of each predictor for model accuracy (although the scale is neither a percentage nor a count of observations). From that, it can be inferred that the variable that contributes the most to the model is the interaction between the normalized match position and the number of words in the IU.

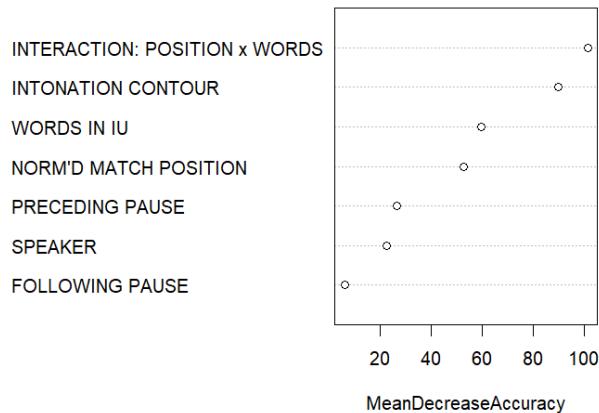


Figure 4: Mean Decrease Accuracy plot from the Random Forest model

To zoom in on the effect of this predictor, I plotted the partial dependence (PD) scores, which represent the dependence between the target response (i.e., copula vs. placeholder) and the different levels of the variable, in this case the product of the normalized match position with the number of words in the IU. That is, PD scores show the average prediction change between copula and placeholder as the chosen variable varies. PD scores marginalize out all of the other features in the model by averaging over predictions based on all combinations of those other features. Given the unbalanced distribution, this process creates a bias towards placeholders, since the combinations of the other features tend to be those that occur in placeholders. Thus, any PD score below 0.67 (i.e., the probability of the placeholder) can be interpreted as favoring the copula, whereas scores above 0.67 can be interpreted as favoring the placeholder (see Hastie et al. 2009: 370 for bias on PD scores). In a nutshell, due to the unbalanced distribution of the two functions (copula vs. placeholder), lower PD scores represent a preference for the copula, whereas higher PD scores represent a preference for the placeholder.

With this in mind, the plotting of the PD scores in Figure 5 shows that the placeholder is predicted for the lower values of the interaction predictor (i.e., 0 to 1), whereas the copula is predicted for higher values (i.e., roughly 1.3 and above). The jittered rugs at the bottom show the distribution of each value of the variable. These results are consistent with the predictions that the copula is more likely to appear towards the end of an IU and in IUs with more elements, and that the placeholder is more likely to appear towards the beginning of an IU and in IUs with fewer elements.

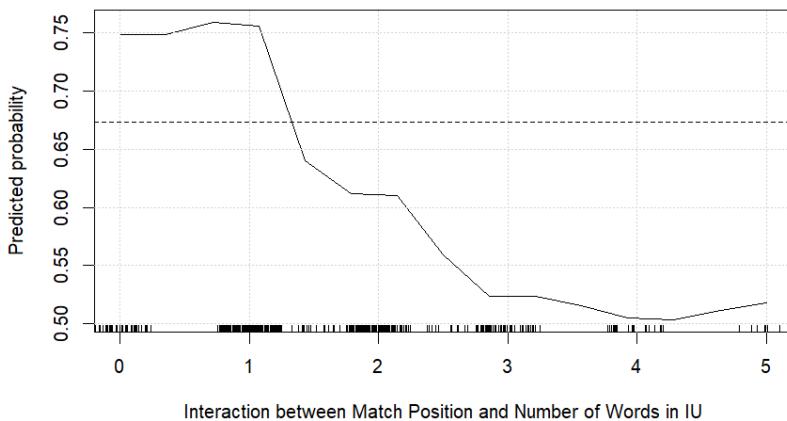


Figure 5: PD scores of the interaction, with dashed line representing the threshold above which the placeholder is predicted; the copula is predicted below the threshold line

Another way to represent the effect of the interaction is to plot the two individual predictors in the interaction (i.e., normalized match position and the number of words in an IU) in a heat map, with the predicted probability in the corresponding colored cell range. Figure 6 shows higher predicted probabilities for the copula towards the top and the right sections, which represents longer IUs and the copula appearing towards the end of an IU. The more purple and blue cells represent lower probabilities for the copula and, thus, higher probabilities for the placeholder. These are mostly found towards the bottom and the left sides of the heat map, suggesting a higher likelihood for the placeholder to appear towards the beginning of an IU and in IUs with fewer elements. Notably, IUs with a single word (which are normalized to a match position of 1) also show a stronger preference for the placeholder.

According to the Mean Decrease Accuracy plot, the second variable that contributes the most to the model is the intonation contour. Figure 7 shows the PD scores for the placeholder as a function of the intonation contour, where bar width represents the proportion of observations in each intonation category and the horizontal dashed line represents the 0.67 threshold. In this case the placeholder is predicted for the continuing and truncated intonation contours, whereas the copula is predicted for the falling intonation. This result is thus consistent with the expectations for each function.

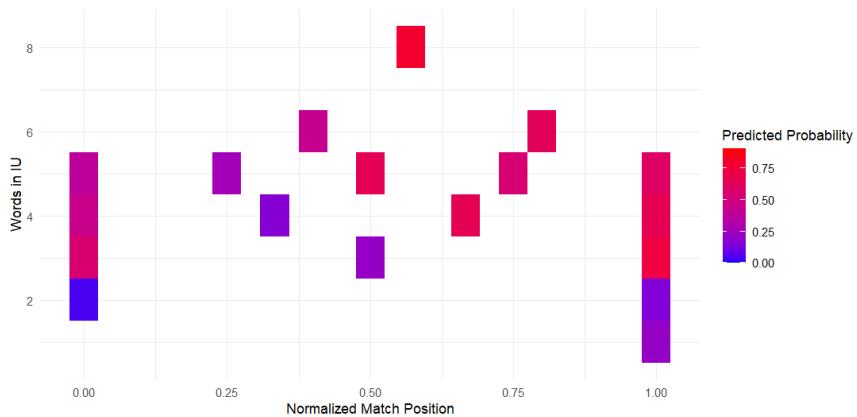


Figure 6: PD scores of the interaction as a function of both normalized match position and number of words in an IU plotted independently. Each cell represents the probability of occurrence of the copula based on the combination of the two variables, with red shades representing higher probabilities. The probabilities for the placeholder are complementary and thus are not displayed here

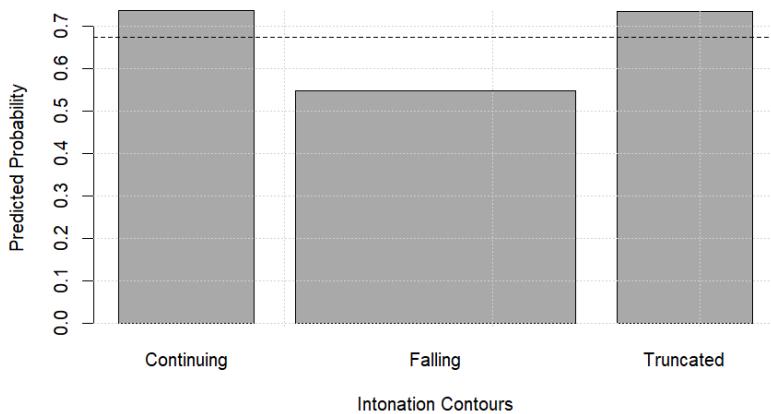


Figure 7: PD scores of the intonation, with the dashed line representing the threshold above which the placeholder is predicted; the copula is predicted below the threshold line

4.4 Discussion

The results from the random forest modeling show that the copula and the placeholder have slightly different distributions in discourse, with the placeholder displaying a stronger tendency towards the beginning of an IU and in IUs with fewer elements, and the copula displaying a tendency towards the end of an IU and in IUs with more elements. Additionally, the model shows a stronger preference for the placeholder with continuing and truncated intonation contours, and a stronger preference for the copula with falling intonation. These results add to the qualitative evidence that these forms are synchronically distinct and suggest that a polysemy analysis is unlikely.

The results for the pause information and the speaker predictors suggest that these variables do not contribute substantially to predicting each function. At first, it might appear somewhat surprising that the preceding pause is not a contributing factor. However, Clark & Fox Tree (2002) show that preceding pause lengths explain the differences between two distinct fillers (i.e., English *uh* and *uhm*), whose main function is to signal a delay (minor for *uh*, major for *uhm*) while holding the floor. In the case of *λe-*, both the copula and the placeholder are morphosyntactically integrated in the discourse in progress and, thus, are not directly comparable to Clark & Fox Tree (2002)'s analysis. These differences in discourse and morphosyntactic environments might help explain why the predicting power of the pause information is limited in this case (but see other papers in this volume).

5 Origins of the placeholder

The outcomes of the qualitative and quantitative analyses now allow us to attempt to answer Epps (2008)'s questions, i.e. whether the distinct functions of *λe-* are a case of polysemy, a historical accident of independent forms converging into the same phonological shape, or a diachronic divergence from the same source. Both functions appear synchronically distinct and identifiable on the prototypical roles of copula and placeholder. Additionally, their different distributions in discourse suggest that a polysemy scenario seems unlikely, under the assumption that related senses of a word display similar distributions. Rather, the two uses of *λe-* are best analyzed as being associated with different morphs. The question about their historical origins, however, remains open, that is, whether their phonological resemblance is a historical accident or whether the two are historically related.

To answer this question the best place to start is by analyzing the historical record of the copula. Nikolaeva (2006, 2020) suggests that the copula *λe-* can be reconstructed in Proto-Yukaghir as **λe-*, since an identical copular form is attested in Tundra Yukaghir. In fact, she argues that the copula is cognate with the Uralic **le-* ‘to exist, to become’ (but see Aikio 2014 for a dissenting argument). Additional evidence for the copula *λe-* being somewhat older material is that at least two suffixes can be traced back to it: the evidential *-λel-* and the predicative case marker *-lek* (Nikolaeva 2020). The evidential arose as a combination of the copula *λe-* with the participle *-l* as a modal construction that was later reanalyzed as inflectional mood. As Nikolaeva (2020) points out, other suffixes in Kolyma Yukaghir reflect a similar development, and the pathway from copula to evidential is attested cross-linguistically (Aikhenvald 2004). As for the predicative case marker, Nikolaeva (2020) argues that *-lek* reflects the combination of the copula *λe-* with the focus marker *-k*, a possible development considering that the alternation between *λ~l* is not always phonologically conditioned in Yukaghirc (Nikolaeva 2020). Combined, *λe-k* served as an existential predicate following a noun. Over time, the copula and the focus marker fused together, but the resulting nominal case marker maintained some of its properties, including the ability to signal assertions without a verbal predicate.

Another avenue for unraveling the historical relationship between the copula and the placeholder is to consider cross-linguistic patterns. As mentioned, demonstratives are a common source for placeholders cross-linguistically (Hayashi & Yoon 2010, Podlesskaya 2010), an unlikely source in Kolyma Yukaghir since the placeholder can target demonstratives. Nikolaeva (2020: 54)’s reconstruction of the demonstrative stems in Proto-Yukaghir confirms this suspicion, as they closely resemble the proximal *tii-*, the medial *adaa-* and the distal *taa-*. However, Nagasaki (2010)’s glossing of the placeholder as “that one” captures the pronominal quality of the placeholder, unlike hesitators (Hayashi & Yoon 2010). This referential flavor is probably amplified when the placeholder is used “approximately” (Podlesskaya 2010) and is deemed as a sufficient referent without later clarification on the target.

Copulas, too, have been shown to grammaticalize from demonstratives (Kuteva et al. 2019), but they are also the source of third person pronouns in what Katz (1996) calls the “copula cycle.” Hebrew is a good example of this cycle: the copula in Proto-Semitic gave rise to an independent third person pronoun in Biblical Hebrew, which in turn evolved to become a new copula in modern Hebrew (Katz 1996). Additional examples of copulas turning into markers of third person include Turkish (Katz 1996), Western Iranian languages (Korn 2011), and Guinea-Bissau Creole (Truppi 2021).

The connection of copulas and placeholders to demonstratives and pronominal forms does not seem accidental. Rather, it is plausible to assume that referentiality is what connects the copula *λe-* to the placeholder *λe-* historically. However, it is hard to posit one function as the source of the other given the copula's cyclic pattern with demonstratives and deictic pronouns. Theoretically, at least, both developments (i.e., from placeholder to copula, and from copula to placeholder) are possible.⁷ It is somewhat of a chicken and egg situation.

The development of the copula from the placeholder is plausible by extending Podlesskaya (2010)'s "approximate naming" function to contexts where the target would be the main assertion. Given *λe-*'s ability to stand in for a target that functions "nominally" and "verbally", it is possible that speakers might not always clarify what their intended target is and deem the placeholder as a sufficient predication device. When following lexical items deployed as converbs, such a structure can be reanalyzed as an analytic construction with an auxiliary. Such constructions are attested in Kolyma Yukaghir (e.g., the so-called periphrastic prospective, Maslova 2003: 178; Nagasaki 2010: 244).

Example (14) shows an instance of a bridging context, where the use of the placeholder could be interpreted also as a copula. The speaker here attaches the so-called verb-focus transitive morphology to *λe-*, which suggests that *λe-* is in fact standing in for a different lexical item; the copula is not otherwise attested with this assertion morpheme. However, the target is not retrieved and the predication is deemed sufficient, maybe because the preceding converb already provides enough lexical information. Without clarification of the target, this structure could be interpreted with the converb as the main lexical contributor to the discourse, where *λe-* has a supporting role to fulfill the need for an assertion. This interpretation can give rise to a light-verb, copular or existential meaning of *λe-*.

- (14) *Хајсааги јуөлүгэ,*
 qanṣaa-gi juø-lu-ge
 pipe-3POSS see-1SG-LOC
 'When I saw his pipe'
 (0.5)
Эл чубукнъэ,
 el tʃubuk-ne
 NEG chibouk-COM
 'Without chibouk'

⁷A third scenario is also possible: both the placeholder and the copula emerged from an older demonstrative now lost.

(1.4)

түдә таŋ чубуккәлә чуму ләндәт,
tude ta-ŋ tʃubuk-kеле tʃumu leŋ-de-t
3SG.GEN that-LNK chibouk-ACC all eat-UNK-CVB.CTX
'Eating all of that chibouk of his'

(1.0)

ЛъЭЛЪЭЛУМ.

λε-λελ-у-м

RH-EV-(EP)-ASS.TR.EF.3SG

'He whatchamacallited'

Potential interpretation: "He had smoked all his chibouk"

(Nikolaeva & Mayer 2004, 34: IU 264–270)

The alternative is to posit the development of the placeholder from the copula. Katz (1996) argues that an intermediate step in the evolution from the copula in Proto-Semitic to the third person pronoun in Biblical Hebrew is its interpretation in a more nominal manner (i.e., 'a/the/his being'). This development is also plausible because of the interpretation of Kolyma Yukaghir roots in more nominal or more verbal ways depending on the context, sometimes without clear signs of derivation (Ventayol-Boada et al. 2023). The nominal interpretation might be part of the reason why *λe-* is glossed as a copula in its placeholder function in Nikolaeva & Mayer (2004), as mentioned previously. The nominal reading of *λe-* is later characterized in more specific terms. Example (15) shows an instance where such an interpretation would be possible. Over time, this structure can be reanalyzed as *λe-* standing in for the characterization, especially if longer delays and unfilled pauses start occurring between the two, and thus giving rise to the placeholder function.

- (15) *Поинэй,*
pojne-j
white-PTCP
'(Which) whites'

(0.8)

нъаасъэнъдъэ,
naače-ŋ-dže
face-PROP-PTCP
'Having a face'

(2.6)

*ЛъЭГЭЛЭ,**λe-gele*

COP-ACC

‘[A] being’

көйгэлэ *йүөнүдэ* моннупльэлни,

køj-gele juø-nide mon-nu-λel-ŋi-Ø

boy-ACC see-CVB.COND say-IMPF-EV-3PL-ASS.INTR.EF

‘If they saw (a being), a boy (with a white face), they said,’

(0.4)

нъаатлэбийэ монут ньуутиэлвэлниа.

naatlebie mon-u-t riuu-t-ie-λel-ŋaa

willow.ptarmigan say-(EP)-CVB.CTX call-UNK-INCH-EV-ASS.TR.EF.3PL

‘They called [him], namely a ptarmigan’

(Nikolaeva & Mayer 2004, 44: IU 90–97)

At first, both scenarios seem plausible: the development of the copula from placeholder, and the development of the placeholder from copula. I believe the latter is more likely in the case of Kolyma Yukaghir for two reasons. First, the copula *λe-* appears to be older linguistic material, as exemplified above by at least two suffixes having likely evolved from it. The fact that an identical copula exists in Tundra Yukaghir adds to this evidence. And second, the placeholder is followed by its target in 71.8% of the time, as mentioned in §4.1. It is unclear whether the frequency of “approximate naming” placeholders as a predicating device would constitute enough grounds to trigger the development of the copula, especially since the remaining 28.2% include both nominal and verbal targets. Table 3 shows the potential stages of the placeholder development from the copula in Kolyma Yukaghir.

Table 3: Hypothetical copula-to-placeholder development in Kolyma Yukaghir

Proto-Yukaghir	>	Kolyma Yukaghir
<i>λe-</i>		<i>λe-</i>
“to be” (cop)	“a/the/his being”	“whatchamacallit” (PH)

Additional evidence for the placeholder development from the copula in Kolyma Yukaghir might be found in Tundra Yukaghir. Given that both languages share the copula *λe-*, it is possible to analyze its distribution in Tundra Yukaghir discourse. The lack of a placeholder use of *λe-* or the presence of a placeholder with a different phonological shape would support the development from copula to placeholder in Kolyma Yukaghir. Alternatively, if *λe-* is also attested as a placeholder in Tundra Yukaghir, the idea that the placeholder came first, from which the copula evolved, would still be possible, although a parallel development in both languages from copula to placeholder as outlined above could not be ruled out either. These questions remain for further research.

6 Conclusion

Analyzing grammatical morphs with identical phonological forms and distinct functions is a common linguistic challenge. In this context, at least three scenarios are possible: 1) the resemblance is a historical accident of two independent elements converging in form, 2) the different functions represent an example of polysemy (i.e., a single form with vaguely related meanings), or 3) the forms are synchronically distinct but historically related (Epps 2008). This Chapter addresses the importance of considering discourse-level explanations when attempting to choose among these scenarios.

The case of *λe-* in Kolyma Yukaghir shows that dividing spoken discourse into IUs can help analyze the semantic and pragmatic components of a form and thus be more confident of its function(s). When the same phonological material appears to be associated with multiple functions, as in the case of *λe-*, discourse features offer the possibility to quantify and model what features are associated more closely with each function, thus allowing us to establish whether their distributions are distinct and rule out a polysemy analysis. Establishing the historical relationship between both morphs can be then attempted by identifying the bridging contexts in discourse, where multiple interpretations of a form are possible and the mechanisms of language change are set in motion.

I argue that the description of *λe-* as a copula in Kolyma Yukaghir fails to account for a significant number of occurrences in naturalistic discourse, which can be better analyzed as a placeholder. Based on their distributions, I conclude that these two functions of *λe-* can be best described as distinct morphs rather than a single polysemous morph with meanings vaguely related. The copula appears to be more likely to appear towards the end of an IU and in IUs with more elements, whereas the placeholder is more likely to occur towards the beginning

of an IU and in IUs with fewer elements. Their resemblance, however, does not appear to be a historical accident but rather connected to referentiality. It seems plausible that the polycategorial nature of Kolyma Yukaghir roots has allowed a nominal interpretation of the copula *λe-* (i.e., ‘a being’) as a referential expression, a development attested cross-linguistically (Katz 1996). When followed by another referential expression that characterizes or delimits it in scope, it creates a morphosyntactic environment that can be reanalyzed as the first element standing in for the second, thus giving rise to the placeholder interpretation.

Overall, this Chapter contributes to the study of discourse and grammar, and the grammaticalization of salient discourse patterns in particular (Ariel 2009, Couper-Kuhlen & Thompson 2008, Du Bois 2003). Additionally, it adds to the growing research on fillers and the emergence of placeholders cross-linguistically by establishing the copula as a new potential source. It does so by offering different methodological tools that can be easily applied in the context of endangered language description.

Abbreviations

(EP)	epenthesis	DIM	diminutive	PF	participant-focus
1	first person	DP	discourse particle	PFV	perfective
3	third person	EF	event-focus	PH	placeholder
ABL	ablative	EV	evidential	PL	plural
ACC	accusative	FUT	future	POSS	possessive
ADV	adverbial	GEN	genitive	PRED	predicative
ASS	assertion	HAB	habitual	PROP	proprietary
AUG	augmentative	IMPF	imperfective	PTCP	participle
CAUS	causative	INCH	inchoative	PURP	purposive
COM	comitative	INS	instrumental	RES	resultative
COND	conditional	INTR	intransitive	SEQ	sequential
COP	copula	LNK	linker	SG	singular
CTX	contextual	LOC	locative	TR	transitive
CVB	converb	NEG	negative	UNK	unknown/unclear
DAT	dative	NMLZ	nominalizer		

References

- Abondolo, Daniel. 1998. Introduction. In Daniel Abondolo (ed.), *The Uralic languages*, 1–42. London: Routledge.

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Aikio, Ante. 2014. The Uralic-Yukaghir lexical correspondences: Genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62. 7–76.
- Anderson, Gregory D. S. 2006. Towards a typology of the Siberian linguistic area. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 266–300. London: Palgrave Macmillan. DOI: 10.1057/9780230287617.
- Anderson, Gregory D. S. 2017. The changing profile of case marking in the north-eastern Siberia area. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 627–650. Cambridge: Cambridge University Press. DOI: 10.1017/9781107279872.024.
- Ariel, Mira. 2009. Discourse, grammar, discourse. *Discourse Studies* 11(1). 5–36. DOI: 10.1177/1461445608098496.
- Baayen, R. Harald. 2011. Corpus linguistics and naïve discriminative learning. *Brazilian Journal of Applied Linguistics* 2(11). 295–328.
- Breiman, Leo. 2001. Random forests. *Machine Learning* 1(54). 5–32.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chafe, Wallace L. 1979. The flow of thought and the flow of language. In Talmy Givón (ed.), *Discourse and syntax*, 159–181. New York: Brill. DOI: 10.1163/9789004368897_008.
- Chafe, Wallace L. 1992. Intonation units and prominences in English natural discourse. In *Proceedings of the IRCS Workshop on Prosody in Natural Speech*, 41–52. Philadelphia: University of Pennsylvania Press.
- Chafe, Wallace L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Citko, Barbara. 2014. Copula. In Mark Aronoff (ed.), *Oxford bibliographies online - linguistics*. Oxford: Oxford University Press. DOI: 10.1093/obo/9780199772810-0196.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Comrie, Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Couper-Kuhnlen, Elizabeth & Sandra A. Thompson. 2008. On assessing situations and events in conversation: ‘extraposition’ and its relatives. *Discourse Studies* 10(4). 443–467.

- Davis, Joseph. 2017. The semantic difference between Italian *vi* and *ci*. *Lingua* 200. 107–121. DOI: 10.1016/j.lingua.2017.08.007.
- Diver, William. 1982. The focus-control interlock in Latin. *Columbia University Working Papers in Linguistics* 7. 13–31.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855. DOI: 10.2307/415719.
- Du Bois, John W. 2003. Discourse and grammar. In Michael Tomasello (ed.), *The new psychology of language. Cognitive and functional approaches to language structure*, vol. 2, 47–87. London: Lawrence Erlbaum Associates Publishers.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane A. Edwards & Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale: Lawrence Erlbaum Associates Publishers.
- Epps, Patience. 2008. From ‘wood’ to future tense: Nominal origins of the future construction in Hup. *Studies in Language* 32(2). 383–404. DOI: 10.1075/sl.32.2.05epp.
- Forina, Michele, Monica Casale, Paolo Olivero & Silvia Lanteri. 2009. CAIMAN brothers: A family of powerful classification and class modeling techniques. *Chemometrics and Intelligent Laboratory Systems* 2(96). 239–245.
- Fortescue, Michael D. 1998. *Language relations across Bering Strait: Reappraising the archaeological and linguistic evidence*. London: Cassell. DOI: 10.1086/466422.
- Fox, Barbara A. & Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66(2). 297–316.
- Gogolev, Anatolij I., Valentina S. Akimova & Vjačeslav I. Šadrin. 2022. Ètnogenez jukagirov: Obzor sovremennoogo sostojanija rešenija problemy [Ethnogenesis of the Yukaghirs: Overview of the current state of the solution to the problem]. *Voprosy nacional'nych i federalivnyx otnošenij* 12(93). 4405–4415.
- Grebenjuk, Pavel S. & Alexander J. Fedorčenko. 2018. Jukagirskaja problema v sovremennoj istoriografii [The Yukaghirs problem in modern historiography]. In *Čelovek i sever. Antropologija, arxeologija, èkologija: Materialy vserossijskoj naučnoj konferencii*, 331–334. Tjumen': Tjumenskij naučnyj centr S.O.R.A.N.
- Grenoble, Lenore A. 2000. Morphosyntactic change: The impact of Russian on Evenki. *Studies in Slavic and General Linguistics* 28. 105–120.
- Grenoble, Lenore A. 2003. *Language policy in the Soviet Union*. New York: Kluwer Academic Publishers.
- Grenoble, Lenore A. 2022. Contact and shift: Colonization and urbanization in the Arctic. In Salikoko S. Mufwene & Ana María Escobar (eds.), *The Cambridge handbook of language contact*, vol. 2: Multilingualism in Population Structure,

- 473–501. Cambridge: Cambridge University Press. DOI: 10.1017/9781009105965. 024.
- Gries, Stefan Th. 2021. *Statistics for linguistics with R: A practical introduction*. 3rd edition. Berlin: De Gruyter Mouton.
- Hastie, Trevor, Robert Tibshirani & Jerome Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. 2nd edition. New York: Springer.
- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders*, 33–66. Amsterdam: John Benjamins.
- Himmelmann, Nikolaus P. 2022. Prosodic phrasing and the emergence of phrase structure. *Linguistics* 60(3). 715–743. DOI: 10.1515/ling-2020-0135.
- Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245. DOI: 10.1017/S0952675718000039.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299. DOI: 10.2307/413757.
- Jochelson, Waldemar. 1926. *The Yukaghirs and the Yukaghirisized Tungus*. New York: G. E. Strecher. DOI: 10.1525/aa.1929.31.1.02a00200.
- Kantarovich, Jessica. 2020. *Argument structure in language shift: Morphosyntactic variation and grammatical resilience in modern Chukchi*. University of Chicago. (PhD dissertation).
- Katz, Aya. 1996. *Cyclical grammaticalization and the cognitive link between pronoun and copula*. Rice University. (Doctoral dissertation).
- Kazakevich, Olga A., Elena M. Budyanskaya, Anastasia P. Evstigneeva, Yuri B. Koryakov, Daria D. Mordashova, Sofie V. Pokrovskaya, Konstantin K. Polivanov, Evgeniya A. Renkovskaya, Zaira M. Khalilova & Karina O. Sheifer. 2022. Škaly jazykovoj vital'nosti i ix primenimost' k materialu konkretnyx jazykovyx situacij [Language vitality scales and their applicability to specific language situations]. *Voprosy Jazykoznanija* 4. 7–47. DOI: 10.31857/0373-658X. 2022.4.7-47.
- Klyachko, Elena. 2025. Placeholders versus general extenders in Tungusic languages. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 41–74. Berlin: Language Science Press. DOI: ??.
- Korn, Agnes. 2011. Pronouns as verbs, verbs as pronouns: Demonstratives and the copula in Iranian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in Iranian linguistics*, 53–70. Wiesbaden: Reichert Verlag.

- Krejnović, Eruxim A. 1982. *Issledovanija i materialy po jukagirskomu jazyku*. Moscow: Akademiia Nauk SSSR.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long & Heiko Narrog. 2019. *World lexicon of grammaticalization*. 2nd edition. Cambridge: Cambridge University Press.
- LeSourd, Philip S. 2003. The noun substitute in Maliseet-Passamaquoddy. In Blair A. Rudes & David C. Costa (eds.), *Essays in Algonquian and Siouan linguistics in memory of Frank T. Siebert, Jr*, 141–163. Winnipeg: University of Manitoba Press.
- Maslova, Elena S. 1997. Yukagir focus in a typological perspective. *Journal of Pragmatics* 27(4). 457–475. DOI: 10.1016/s0378-2166(96)00038-0.
- Maslova, Elena S. (ed.). 2001. *Yukaghirs texts*. Wiesbaden: Harrassowitz Verlag.
- Maslova, Elena S. 2003. *A grammar of Kolyma Yukaghirs*. Berlin: Mouton de Gruyter.
- Matić, Dejan. 2008. Russian influence on the Kolyma Yukagir morphosyntax. In Antje Casaretto & Silvia Kutscher (eds.), *Sprachkontakt, synchron und diachron*, 97–128. Aachen: Shaker.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- Mithun, Marianne. 2008. The extension of dependency beyond the sentence. *Language* 84(1). 69–119. DOI: 10.1353/lan.2008.0054.
- Nagasaki, Iku. 2010. Kolyma Yukaghirs. In Yasuhiro Yamakoshi (ed.), *Grammatical sketches from the field*, 213–256. Tokyo: Research Institute for Languages, Cultures of Asia & Africa (ILCAA).
- Nagasaki, Iku. 2018. The focus construction in early modern Kolyma Yukaghirs. *Gengo Kenkyū (Journal of the Linguistic Society of Japan)* 154(C). 123–152. DOI: 10.11435/gengo.154.0_123.
- Nikolaeva, Irina. 1988. *Problema uralo-jukagirskix genetičeskix svjazej [The problem of the Uralo-Yukaghirs genetic relationship]*. Moscow Institute of Linguistics. (Doctoral dissertation).
- Nikolaeva, Irina. 2006. *A historical dictionary of Yukaghirs*. Berlin: Mouton de Gruyter.
- Nikolaeva, Irina. 2008. Chuvan and Omok languages? In Alexander Lubotsky, Jos Schaeken, Jeroen Wiedenhof, Rick Derksen & Sjoerd Siebinga (eds.), *Evidence and counter-evidence: Essays in honour of Frederik Kortlandt*, vol. 2, 313–336. New York: Brill. DOI: 10.1163/9789401206365_021.
- Nikolaeva, Irina. 2020. *Yukaghirs morphology in a historical and comparative perspective*. Munich: Lincom GmbH.

- Nikolaeva, Irina & Thomas Mayer. 2004. *Online documentation of Kolyma Yukaghir*. <https://www.sgr.fi/yukaghirk/index.html> (10 December, 2021).
- Nikolaeva, Irina & Evgenij Xelimskij. 1997. Jukagirskij jazyk [The Yukaghir language]. In Aleksandr P. Volodin (ed.), *Jazyki mira. Paleoaziatskie jazyki*, 155–168. Moscow: Izdatel'stvo "Indrik".
- Ono, Tsuyoshi, Sandra A. Thompson & Ryoko Suzuki. 2000. The pragmatic nature of the so-called subject marker *ga* in Japanese: Evidence from conversation. *Discourse Studies* 2(1). 55–84. DOI: 10.1177/1461445600002001003.
- Pakendorf, Brigitte. 2024. Language contact in northern Asia. In Edward J. Vajda (ed.), *The languages and linguistics of northern Asia*, 1175–1210. Berlin: De Gruyter Mouton.
- Pakendorf, Brigitte. 2025. A corpus-based investigation of fillers in Negidal (Northern Tungusic). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 75–129. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697579.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Podlesskaya, Vera I. & Andrej A. Kibrik. 2006. Korrekcija ustnoj russkoj monologičeskoj reči po dannym korpusnogo issledovaniya [Correction in Russian spoken monologues: A corpus study]. *Russkij jazyk v naučnom osveščenii* 2(12). 7–55.
- Ponsonnet, Maïa. 2025. Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 315–340. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697591.
- Sankoff, Gillian & Penelope Brown. 1976. The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language* 52(3). 631–666.
- Silverstein, Michael. 1976. Shifters, linguistic categories, and cultural description. In Keith H. Basso & Henry A. Selby (eds.), *Meaning in anthropology*, 11–55. Albuquerque: University of New Mexico Press.
- Skorik, Pjotr (ed.). 1986. *Paleoaziatskie jazyki*. Leningrad: Izdatel'stvo "Nauka".
- Smith, Linda Tuhiwai. 1999. *Decolonizing methodologies: Research and Indigenous peoples*. London: Zed Books Ltd.
- Stern, Nancy. 2006. Tell me about yourself: A unified account of English *-self* pronouns. In Joseph Davis, Radmila J. Gorup & Nancy Stern (eds.), *Advances in functional linguistics: Columbia School beyond its origins*, 177–194. Amsterdam: John Benjamins. DOI: 10.1075/sfsl.57.14ste.

- Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. DOI: 10.1037/a0016973.
- Thompson, Sandra A. & Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 27–60. Amsterdam: John Benjamins.
- Thompson, Sandra A. & Anthony Mulac. 1991. A quantitative perspective on the grammaticalization of epistemic parentheticals in English. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2: Focus on types of grammatical markers, 313–329. Amsterdam: John Benjamins.
- Troiani, Giorgia. 2023. *Representing a language in use: Corpus construction, prosody, and grammar in Kazakh*. University of California, Santa Barbara. (PhD dissertation). <https://escholarship.org/uc/item/9n65973f>.
- Truppi, Chiara. 2021. The copula cycle in Kriol. *Journal of Historical Syntax* 5(1–13). 1–27. DOI: 10.18148/hs/2021.v5i1-13.42.
- Vajda, Edward J. 2009. The languages of Siberia. *Linguistics and Language Compass* 3(1). 424–440. DOI: 10.1111/j.1749-818X.2008.00110.x.
- van Valin, Robert D. Jr. 2003. Functional linguistics. In Mark Aronoff & Janie Rees-Miller (eds.), *The handbook of linguistics*, 319–336. Oxford: Blackwell.
- Vatanen, Anna. 2014. *Responding in overlap: Agency, epistemicity and social action in conversation*. University of Helsinki. (PhD dissertation). <https://helda.helsinki.fi/handle/10138/136165>.
- Ventayol-Boada, Albert. 2023. Applications of classification trees for endangered language description: Finite verb morphology in Kolyma Yukaghir. In *Proceedings of the 6th workshop on the use of computational methods in the study of endangered languages*, 7–12. Remote: Association for Computational Linguistics.
- Ventayol-Boada, Albert, Nathan Roll & Simon Todd. 2023. Unsupervised part-of-speech induction for language description: Modeling documentation materials in Kolyma Yukaghir. In *Proceedings of 2nd workshop on NLP applications to field linguistics*. Dubrovnik: Association for Computational Linguistics.
- Vinokurova, Antonina, Irena Khokholova, Boris Osipov, Stepan Pavlov, Yana Tokhtobina & Vilyuyana Platonova. 2022. The language situation in the Eveno-Bytantaiskii national district of the Republic of Sakha (Yakutia). *Sibirica* 21(3). 159–194. DOI: 10.3167/SIB.2022.210308.

- Visser, Eline. 2025. Form and frequency of Kalamang fillers. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 251–273. Berlin: Language Science Press. DOI: ??.
- Volodin, Aleksandr P. (ed.). 1997. *Jazyki mira. Paleoaziatskie jazyki [The world's languages. The Paleoasiatic languages]*. Moscow: Izdatel'stvo "Indrik".
- Willerslev, Rane. 2004. Not animal, not not-animal: Hunting, imitation and empathetic knowledge among the Siberian Yukaghirs. *Journal of the Royal Anthropological Institute* 10(1). 629–652. DOI: 10.1111/j.1467-9655.2004.00205.x.
- Wouk, Fay. 2005. The syntax of repair in Indonesian. *Discourse Studies* 7(2). 237–258. DOI: 10.1177/1461445605050368.
- Zhao, Yuan & Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. In *Proceedings of DiSS'05: Disfluency in spontaneous speech workshop*, 179–182. Aix-en-Provence: Université de Provence.

Chapter 5

Choosing fillers in Besemah

Bradley McDonnell^a & Blaine Billings^a

^aUniversity of Hawai‘i at Mānoa

Studies of fillers have shown that while some languages, such as English, express (interjective) hesitator and placeholder functions by different means, other languages, such as Chinese, Japanese, and Korean, express these functions with the same word (or class of words). Against this backdrop, the Besemah dialect of South Barisan Malay presents an interesting case study because, like Chinese, Japanese, and Korean, Besemah does not make a formal distinction between hesitator and placeholder functions, meaning one and the same form may serve both functions. However, Besemah differs from these languages in that, despite having a dedicated filler, *anu*, Besemah speakers make extensive use of the proximal demonstrative pronoun *ini* – and to a lesser extent the distal demonstrative pronoun *itu* – with the same range of functions as *anu*. In analyzing these fillers in Besemah, we question some of the defining characteristics of fillers in general, including the strict separation of hesitator and placeholder functions and the degree to which repair is characteristic of placeholders.

keywords: South Barisan Malay, hesitator, placeholder, repair, demonstrative

1 Introduction

In their seminal paper on fillers, Hayashi & Yoon (2006) provide defining characteristics of two different types of fillers: placeholders and (interjective) hesitators. They point out that, in languages like English, there are separate means of fulfilling these two different roles (i.e., the lexicalized phrase placeholder *whatchamacallit* and hesitators *uh* or *uhm*). In other languages like Korean, Japanese, or Chinese, these roles are fulfilled by a single lexical item or a class of lexical items, namely demonstratives. Fillers of both types are defined as conventionalized forms, which deal with constraints in cognitive processes that are associated



with trouble in word formulation as a means of delaying the next word or phrase (see Hayashi 2003, Hayashi & Yoon 2006, Fox 2010). They can be used in similar interactional contexts, such as word searches, different types of repair sequences, and holding or ceding the floor (Clark & Tree 2002). Despite their use in similar contexts with similar purposes, Hayashi & Yoon (2006) define clear differences between the two types of demonstrative fillers, as summarized in Table 1.

Table 1: Defining characteristics of placeholder and hesitator according to Hayashi & Yoon (2006).

Placeholder	Hesitator
“...participates in the morpho-syntactic structure of an unfolding utterance.”	“...are not produced as a syntactic constituent occupying a specific syntactic slot in an ongoing utterance.”
“...is used as a referential expression.”	“...used non-referentially, without making any referential contribution.”
“...is subsequently replaced by a more specific lexical item that has become available to the speaker (and/or hearer(s)) as a result of word search.”	“...little syntactic or semantic correspondence between the morphological forms ...and the ‘outcome’ of word search.”

Essentially, Hayashi & Yoon (2006) distinguish placeholders from hesitators based on their referentiality (i.e., placeholders are referential, while hesitators are non-referential), syntactic integration in the ongoing utterance (i.e., placeholders are syntactically integrated, while hesitators are not), and repair of the filler (i.e., placeholders are “replaced” by a more specific word, whereas the hesitator and the “outcome” have little correspondence).

Against this backdrop, Besemah, a dialect of South Barisan Malay, provides an interesting case study because like Chinese, Japanese, and Korean, Besemah does not make a formal distinction between hesitator and placeholder functions, meaning one and the same form may serve both functions. However, fillers in Besemah differ from these languages in that, despite having a dedicated, identifiable filler, *anu*, Besemah speakers make extensive use of the proximal demonstrative pronoun *ini* – and to a lesser extent the distal demonstrative pronoun *itu* – with the same range of functions as the filler *anu*. As this paper demonstrates, these fillers also share many of the same morphosyntactic properties, distributions, and frequencies.

In describing Besemah fillers, we problematize the notion that there are categorical differences between placeholders and hesitators as defined by Hayashi & Yoon (2006). First, while clear examples of both functions can be identified in the corpus, in a number of cases the syntactic integration of the filler – a distinguishing property of placeholders – is indeterminate. For both analysts and participants, it is often impossible to decide whether a filler is employed as a hesitator or placeholder. Second, we find that the nature of repair is similarly not clear-cut. In Besemah, apparent placeholders are commonly not replaced, and even when they are replaced by a more specific lexical item, the two do not necessarily share the same morphology. Drawing on a sizeable documentary corpus of everyday conversations, we demonstrate how Besemah complicates the typology of fillers in the world’s languages.

This chapter is organized as follows. §2 introduces Besemah, the basic properties of its morphosyntax, and the corpus used in this study. §3 discusses the etymologies of both the dedicated filler *anu* in Besemah and the demonstrative pronouns as fillers before describing how speakers utilize them in interaction. It is further demonstrated how these fillers are used variably as placeholders and hesitators and, in a number of cases, how their role remains indeterminate. The section concludes with a discussion of the frequency counts of different morphological configurations of the fillers as found in the corpus. §4 presents a quantitative analysis of Besemah fillers in a subset of the corpus, describing the frequency of the different fillers (i.e., *anu* versus demonstrative pronouns), their morphosyntactic properties, uses as hesitator or placeholder, properties related to disfluencies and repair, and distributions across speakers. §5 concludes the chapter.

2 Besemah

Besemah is a Malayic language spoken by approximately 330,000 people in the highlands of southwest Sumatra. It is considered a variety of South Barisan Malay, which comprises a complex network of 11 other named dialects spoken by an estimated 1.5 million people. Despite the relatively large number of speakers, South Barisan Malay remains under-described (see McDonnell 2016). Due to their diversity along several dimensions, Malayic languages have been categorized along sociolinguistic lines as vernacular Malay (i.e., varieties regularly inherited from Proto Malayic), literary Malay (i.e., varieties arising from the literary Malay tradition), and vehicular Malay (i.e., ‘trade’ varieties that came about as a result of contact; for an overview of the complexities of these issues see Ade-laar 2005, McDonnell et al. 2024). Unlike some well-known varieties of Malay,

such as Standard Indonesian and Baba Malay, which fall into literary Malay and vehicular Malay categories, respectively, South Barisan Malay is considered a vernacular Malay variety, since these dialects appear to have been regularly inherited from Proto Malayic.

In the remainder of this section, we provide a brief overview of the basic structural properties of Besemah as a foundation for understanding the complexities of fillers in the language. We then follow in §3 with a description of filler use, including competing forms of the fillers and their hesitator and placeholder functions, the use of fillers in trouble in word formulation, the morphological possibilities of fillers, and the indeterminacy of hesitator and placeholder functions. The examples and the data on which the analysis is based in this section come from a corpus of everyday conversations in Besemah that was compiled by the first author. The corpus contains 27 conversations, each involving two to five Besemah speakers. Sessions are each approximately one hour in length and have been transcribed using a simplified version of Discourse Transcription (Du Bois et al. 1993).¹ The recordings are archived with the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC; McDonnell 2008). The conversational corpus contains just over 250,000 words and just under 95,000 Intonation Units (IU).² The quantitative analysis in §4 draws on a subset of this corpus.

2.1 Basic morphosyntactic properties

Besemah is a mildly synthetic language, with a fairly robust set of nominal and verbal prefixes/proclitics and suffixes/enclitics in addition to a few circumfixes. The most common affixes include verbal prefixes that express voice and verbal suffixes that represent polyfunctional, causative/applicative markers. Nominal affixes derive nouns with various meanings (e.g., locative, objective). There are also a number of pronominal clitics that serve both as verbal arguments and possessors (see below). Table 2 provides a brief overview of some of the more

¹Transcription conventions are as follows: Segments are chunked into Intonation Units (IU) represented by a line break. End marks represent a ‘final’ contour (‘.’), a ‘continuing’ contour (‘,’), or an ‘appeal’ contour (‘?’). Dashes represent truncation: en dash (‘–’) for a word and em dash (‘—’) for an IU. Square brackets represent overlapping speech. Colons represent lengthening. Short pauses are represented by ‘..’, while timed long pauses are represented as seconds within parentheses.

²The following sessions are included: BJM01-001, BJM01-002, BJM01-004, BJM01-008, BJM01-010, BJM01-011, BJM01-015, BJM01-028, BJM01-029, BJM01-032, BJM01-033, BJM01-035, BJM01-041, BJM01-043, BJM01-071, BJM01-073, BJM01-081, BJM01-086, BJM01-088, BJM01-093, BJM01-098, BJM01-116, BJM01-125, BJM01-129, BJM01-130, BJM01-131, BJM01-158.

productive morphemes that occur on fillers, excluding morphology that marks transitive verbal predicates (presented later in this section). These affixes readily combine with fillers, as discussed further in §3. For a full listing of Besemah morphology, see McDonnell (2016: 36–45).

Table 2: Several productive affixes in Besemah.

Affix	Meanings with examples
<i>be-</i>	Middle voice (e.g., <i>jalan</i> ‘path’ → <i>bejalan</i> ‘walk’) that also often derives denominal intransitive verbs with meaning of ‘have NOUN’ (e.g., <i>beghas</i> ‘unhusked rice’ → <i>bebeghas</i> ‘have unhusked rice’) or ‘use NOUN’ (e.g., <i>pisau</i> ‘knife’ → <i>bepisau</i> ‘use a knife’) with various other semantics associated with the root (e.g., <i>kerite</i> ‘bike’ → <i>be-kerite</i> ‘ride a bike’)
<i>te-</i>	Non-volitional voice (e.g., <i>teghingat</i> ‘remember’) that also derives intransitive verbs with accidental meaning (e.g., <i>beli</i> ‘buy’ → <i>tebeli</i> ‘inadvertently buy’)
<i>-an</i>	Objective nominalizer commonly derives nouns with meanings of the resulting action of a root (e.g., <i>pecah</i> ‘break’ → <i>pecahan</i> ‘a broken piece’) or the patient associated with the root (e.g., <i>buat</i> ‘make’ → <i>buatan</i> ‘something made’), although it also may result in the location where the root is carried out (e.g., <i>mandi</i> ‘bathe’ → <i>mandian</i> ‘bathing place’) and other conventionalized meanings (e.g., <i>manis</i> ‘sweet’ → <i>manisan</i> ‘desserts’).
<i>se-</i>	Numeral prefix meaning ‘one’ (e.g., <i>se-mobil</i> ‘one car’) may also be used to mean ‘same NOUN’ (e.g., <i>se-mobil</i> ‘(in) the same car’)

Nominal arguments and many intransitive predicates, however, need not occur with any additional morphology, as in the examples in (1).

- (1) a. *die empai duduk*

3 REC.PST sit

‘she just sat down.’

(BJM01-116-01, 01:00:52–01:00:53, Speaker: Susiana)

- b. *dindak duduk die Wi*

not.want sit 3 W.

‘she didn’t want to sit, Wi.’

(BJM01-081-01, 00:13:11–00:13:13, Speaker: Riska)

The examples above demonstrate other properties of Besemah clause structure. Predicates are commonly preceded by auxiliary verbs that express tense, aspect, and/or mood. The single argument, S, may precede, as in (1a), or follow, as in (1b), the predicate.

With only a few exceptions, transitive predicates are marked as either A-Voice (AV) or P-Voice (PV) constructions. In AV, the verb is prefixed with the homorganic nasal N- and the A argument is considered to be syntactically and pragmatically privileged, often functioning as the topic of discourse and able to undergo specific syntactic operations such as relativization. AV is exemplified in (2). In PV, the P argument is considered privileged in the same way, with the additional complication that verbal marking is based upon the person of the A argument. When A is first or second person, it is procliticized to the verb, as in (3), whereas when A is third person, it is encliticized to the verb which is optionally prefixed with *di-*, as in (4). This voice system is said to be symmetrical as both constructions are equally transitive and neither construction appears to be derived from the other (Himmelmann 2005, Riesberg 2014, Chen & McDonnell 2019).

- (2) *aku la udim .. ng-ambik mulan kawe di situ eh*
1SG PFV finish AV-take seed coffee at there TAG
'I already finished taking the coffee seeds there, right.'
(BJM01-001-01, 00:20:50–00:20:53, Speaker: Yawan)
- (3) *nik abang-abang ku=ambik-i*
N.LI red-DISTR 1SG=PV.take-CAUS/LOC
'I took the red ones (i.e. peppers).'
(BJM01-158-01, 00:52:54–00:52:46, Speaker: Neti)
- (4) *se-gale=nye di-tanam=e*
one-all=3 PV-plant=3
'she planted all of them.'
(BJM01-098-01, 00:03:44–00:03:46, Speaker: Ijir)

Grammatical relations are not marked by case or agreement, but S, A in AV, and P in PV pattern together and appear to be the 'privileged syntactic arguments' in the sense of van Valin & LaPolla (1997). McDonnell (2022) refers to this grammatical relation as the *Primary Argument*. The other argument in transitive constructions (i.e., P in AV, A in PV), referred to as the *Secondary Argument*, occurs adjacent to the verbal predicate and forms a single constituent with it. We

refer to this constituent as the *predicate complex*. Primary Arguments occur either before the predicate complex, as in the examples (2)–(4) above, or after the predicate complex, as in (5) for AV and (6) for PV below.

- (5) *m-beli guring-an agi aku*
 AV-buy fry-NMZ again 1SG
 ‘I bought fried snacks again.’

(BJM01-004-01, 00:59:31–00:59:32, Speaker: Peter)

- (6) a. *dide ku=ambik duit tu tadi*
 NEG 1SG=PV.take money DEM.DIST earlier
 ‘I didn’t take the money earlier.’

(BJM01-028-01, 00:13:26–00:13:27, Speaker: Husni)

- b. *la di-ambik=e due: .. pelang*
 PFV PV-take=3 two dike
 ‘he already took two dikes.’

(BJM01-033-01, 00:58:23–00:58:25, Speaker: Mariati)

Polyfunctional applicative suffixes perform a number of functions. The suffix *-ka* prototypically serves benefactive and instrumental applicative functions, and the suffix *-i* prototypically serves locative and goal applicative functions. Both suffixes can additionally perform causative and other functions that derive transitive verbs from various types of intransitive verbal and nominal roots. They are also able to attach to transitive verbal roots with no apparent change in valency, often serving semantic or pragmatic functions (see Truong & McDonnell 2022, McDonnell & Truong 2024 for discussion of the complexities of these suffixes in the languages of western Indonesia). In this paper, we gloss *-ka* as ‘CAUS/BEN’ and *-i* as ‘CAUS/LOC’ to reflect their more prototypical functions.

- (7) *ambik-ka aku ayik Put*
 PV.take-CAUS/BEN 1SG water P.
 ‘grab me some water, Put.’

(BJM01-073-01, 00:43:15–00:43:17, Speaker: Nisa)

- (8) *pinggir ni masih tanam-i ngah padi*
 edge DEM.PROX still PV.plant-CAUS/LOC with rice
 ‘the perimeter was still planted with rice.’

(BJM01-011-01, 00:12:37–00:12:39, Speaker: Jamisah)

Within the NP, numerals precede the head noun and other modifiers follow the head noun. NPs are often followed by a demonstrative determiner *ni* ‘DEM.PROX’ or *tu* ‘DEM.DIST’. Consider both lines in the example in (9).

- (9) 1 L: *ame ikan lele ni*,
 TOP fish catfish DEM.PROX
 ‘as for the catfish,’
 2 *di-untal-ka baih ayam putung mati-mati tu bik*
 PV-toss-CAUS/BEN just chicken cut die-DISTR DEM.DIST aunt
 ‘throw to them butchered chickens that are dead, auntie.’
 (BJM01-125-01, 00:15:23–00:15:26, Speaker: Lis)

Besemah, like many other languages of the region, makes use of minimal structures where predicates occur without explicit arguments (McDonnell 2016). The absence of explicit arguments is most commonly analyzed as a form of ellipsis. However, interactional approaches over the past decade have not made such a claim and do not assume that arguments are elided (see Ewing 2019). Instead, they take a what-you-see-is-what-you-get approach, where minimal structures are not outliers but the norm. Consider the example in (10) below.

- (10) 1 L: *la beghape kali ng-gureng?*
 PFV how.many times AV-fry
 ‘how many times did (you) fry (the fishcakes).’
 2 *tujuh tujuh saje*
 seven seven continuously
 ‘seven, seven every time.’
 3 N: *dide tujuh saje Ci*
 NEG seven continuously C.
 ‘not seven every time, Ci.’
 (0.8)
 4 *dapat beghape?*
 get how.many
 ‘how many (fishcakes) do (we) have?’
 5 *dide pule ng-itung eh*
 NEG also AV-count TAG
 ‘(we) didn’t really count (them), right.’
 (BJM01-158-01, 01:13:17–01:13:22, Speakers: Leksi, Neti)

3 Fillers in Besemah

As mentioned in the introduction, Besemah makes use of two categories of fillers. The first is the dedicated filler *anu*, which is often glossed as ‘whatchamacallit’ or as ‘whatshername’ or ‘whatshisname’. This filler is widespread in Malayic languages and is reconstructed to Proto Malayic *anu? ‘something; someone, so-and-so’ as an indefinite pronoun (Adelaar 1992: 128–129). At higher levels, Blust (2013: 516) reconstructs Proto Malayo-Polynesian *anu as an indefinite interrogative, providing the gloss ‘whatchamacallit’, and, at the top level of the family, Proto Austronesian *anu as a general interrogative, providing the gloss ‘what’. Thus, the source of the filler in Besemah is likely the interrogative pronoun *anu that can be reconstructed all the way back to Proto Austronesian. What is remarkable is that the use of *anu as a filler can be reconstructed all the way back to Proto Malayo-Polynesian and that this form is today used as a filler in a diverse set of Malayo-Polynesian languages spoken throughout the Philippines and Indonesia (see Blust et al. 2023 for a list of reflexes).³

The second category is the demonstrative pronouns *ini*, a proximal demonstrative, and *itu*, a distal demonstrative, although the latter is much less often used as a filler. While these forms are familiar from other Malayic languages and are reconstructed to Proto Malayic (Adelaar 1992: 129), Besemah actually contrasts two forms of the demonstrative pronouns with innovative forms *tini*, for the proximal, and *titu*, for the distal. In prototypical demonstrative (i.e. non-filler) functions, the two sets of forms appear to behave the same, and in discussions with Besemah speakers, the forms are completely interchangeable. However, when functioning as a filler, the original forms *ini* and *itu* are far more common. Short forms of these demonstratives, *ni* and *tu*, may only be used as demonstrative determiners, marking the end of a NP or certain subordinate clauses.

3.1 Placeholder and hesitator functions

Hesitator and placeholder functions are evident for both *anu* and the demonstrative pronouns *ini* ‘DEM.PROX’ and *itu* ‘DEM.DIST’. The examples below demonstrate how each type is used in both functions: (11) and (12) exemplify placeholders for *anu* and *ini*, respectively; (13) and (14) demonstrate hesitator functions for *anu* and *ini*, respectively.

In line 1 of (11), Jamisah is discussing a road on the downriver side of the village, and in line 2, she employs the filler *anu* between an auxiliary, *masih*

³See Nagaya (2022: 92–93) for an alternative analysis, wherein the interrogative pronoun developed from a placeholder.

‘still’, and a PP, *di lembak ni* ‘on the downriver side’. This position is reserved for predicates and, as we will see below, is recognizable as such to participants. Thus, *anu* functions here as a placeholder for a predicate, one which is repaired in line 3 along with the recycled auxiliary *masih* ‘still’. In this and all following multi-line examples, an arrow (→) is used to indicate the line in which the filler is used.

- (11) 1 J: *ame nik di lembak ni*,
 TOP N.LI at downriver DEM.PROX
 ‘as for the downriver one (i.e. road),’
 2 → *masih anu di lembak ni*,
 still FILL at downriver DEM.PROX
 ‘still whatchamacallit on the downriver (side),’
 3 *masih rusak*
 still broken
 ‘(it) is still broken down.’
 (BJM01-011-01, 00:31:06–00:31:09, Speaker: Jamisah)

In the talk just prior to (12), the participants are airing their grievances about certain people passing through their fields, when in line 1 Partiwi seeks to bring the topic to a close by stating that they just need to worry about themselves and their families (i.e. ‘think about our children’). This example is similar to the previous one except that the speaker, Partiwi, employs in line 1 the proximal demonstrative *ini* ‘DEM.PROX’ just after an auxiliary *nak* ‘want’. Again, this position is evidence that the filler is acting as a placeholder. The placeholder is repaired in line 2 with the predicate complex *nginaki dai anak* ‘to think about our children (lit. to look at (our) child’s face)’. What is particularly interesting here is that, while she recycles the auxiliary *nak* ‘want’ in the repair, the placeholder *ini* does not share any of the verbal morphology with what is repaired. We return to this mismatch between the morphology of the placeholder and the repair in §3.2.

- (12) 1 →P: *tape ame diwik nak ini eh*
 what TOP self want DEM.PROX TAG
 ‘because, as for us, (we) should whatchamacallit, right?’
 2 *ka- nak ng-(k)inak-i dai anak*
 TRUN want AV-see-CAUS/LOC face child
 ‘(we) should think of our children (lit. want to look at (our) child’s face),’
 (BJM01-125-01, 00:40:02–00:40:06, Speaker: Partiwi)

Fox (2010: 2), following terminology from Interactional Linguistics, describes placeholders as “fulfill[ing] the syntactic projection of the turn so far.” In the examples above, the placeholders *anu* and *ini* fulfill the syntactic projection of the predicate and predicate complex, respectively; the former is repaired by the predicate *rusak* ‘broken’ in (11) and the latter by the predicate complex *nginaki dai anak* ‘to look at (our) child’s face’ in (12). Podlesskaya (2010) uses the term *target*, which refers to the intended word (or phrase) that the filler is taking the place of. Thus, the predicate *rusak* ‘broken’ in (11) and the predicate complex *nginaki dai anak* ‘to look at (our) child’s face’ in (12) are considered the targets. While notions of syntactic projection and target are conceptually different, in practice they can be used interchangeably when describing the syntactic properties of placeholders. However, it is important to note that these notions often assume some sort of repair (i.e., a replacement of the placeholder) or, at the very least that there is a particular, identifiable intended word that the placeholder takes the place of. As we show in §4.5, the majority of placeholders in Besemah are not repaired and in a number of cases it is not entirely clear what the intended word is or that there is even an intended word at all. For this reason, we simply describe the projection that the placeholder fulfills or simply the syntactic position in which the placeholder occurs (see §4.4).

Turning to the hesitator functions of the same fillers, the example in (13) demonstrates how *anu* is not syntactically integrated into the utterance and thus not recognizable as a placeholder. In line 1, Dewi is explaining that really muddy rice paddies are nice to work in. After a short pause, she adds the qualifier in line 2. She uses the filler *anu* before *sambil* ‘while’, which marks simultaneous adverbial clauses. In the syntax of these simultaneous clauses, any elements occurring before *sambil* ‘while’ are not syntactically integrated into the clause. Thus, the filler here is analyzed as a hesitator.

- (13) 1 D: *anye ame kampung lacak tu lemak die*
 but TOP village mud DEM.DIST pleasant 3
 ‘but, if (it’s) muddy, it’s nice.’
 (0.4)
- 2 → *anu sambil mundur*
 FILL while retreat
 ‘uhm while going backwards.’
 (BJM01-011-01, 00:12:16–00:12:20, Speaker: Dewi)

In (14), Neti and Leksi are talking about how Leksi’s husband fixed his health insurance card. Neti summarizes what Leksi had said in line 1, and then, after this

fully articulated transitive clause, she uses the proximal demonstrative *ini* as a filler in its own IU in line 2. The filler is followed by a pause and a self-addressed question before Neti asks her question to Leksi in the final line. This use of the filler is again not integrated into the clause and acts on its own as a hesitator.

- (14) 1 N: *Pinsi ng-iluk-i* KIS *tu*⁴
P AV-beautiful-CAUS/LOC medical.card DEM.DIST
'Pinsi fixed the medical card.'
(0.6)
- 2 → *ini*,
DEM.PROX
'uhm,'
(0.4)
- 3 *tape?*
what
'what?'
(0.7)
- 4 *langsung s(e)=aghī Ci?*
direct one=day C.
'in a single day, Ci?'

(BJM01-158-01, 00:35:21–00:35:27, Speaker: Neti)

When either form of the filler takes affix or clitic morphology, we analyze the filler as performing a placeholder function, fitting with Hayashi & Yoon's (2006) criterion of morphosyntactic integration (see §1). While the observed morphological forms of placeholders are discussed in the next subsection, the following examples demonstrate the placeholder functions of affixed fillers.

In (15), Royani is telling her granddaughter that she should not throw away her rice. She uses *anu* in line 2 with a third-person possessive enclitic *=nye*. Aside from the enclitic marking, *anu* is followed by the demonstrative determiner *tu* and follows the PV-marked predicate, both of which provide further support for the syntactic integration of the filler (i.e., as a noun). Note that the enclitic *=nye* in Besemah has a number of functions other than indicating third person possession and A arguments in PV constructions, including marking for definiteness. Similar uses of the third singular pronoun are found in Nasal (see Billings & McDonnell 2025 [this volume]).

⁴KIS stands for *Kartu Indonesia Sehat* 'Healthy Indonesia Card' which is used to access free health services.

- (15) 1 R: *janga:n*
 NEG.IMP
 ‘don’t.’
 2 → *capak-i* *anu=nye tu,*
 PV.discard-CAUS/LOC FILL=3 DEM.DIST
 ‘throw away her whatchamacallit,’
 3 *nasik=(ny)e*
 rice=3
 ‘her rice.’
 (BJM01-028-01, 00:34:01–00:34:04, Speaker: Royani)

The same sort of situation is found with the demonstrative filler *ini* in (16). In this conversation, participants are weaving, and in what immediately precedes (16), they are discussing how they go into the jungle to find particular plants that are good for weaving. In line 1, Riska uses the filler with the enclitic *=nye* when stating that something is all gone. After a self-addressed question, the filler is repaired in line 3. Aside from the enclitic marking, the post-predicate position is recognizable as a position for Primary Arguments, providing support for a placeholder analysis.

- (16) 1 →R: *la abis die ini=nye*
 PFV finish INTS DEM.PROX=3
 ‘the whatchamacallit was all gone.’
 2 *tape,*
 what
 ‘what,’
 3 *bemban=(ny)e*
 plant=3
 ‘the *bemban* plant.’
 (BJM01-081-01, 00:21:16–00:21:18, Speaker: Riska)

As we will see further below, if placeholders occur with morphology, the morphology is most frequently verbal. In the example in (17), the filler in line 1 is marked with the so-called middle voice prefix *be-* (see Table 2 for a description of this prefix). Its position between the negator *dide* ‘NEG’ and the adverb *agi* ‘again’ is a recognizable predicate position. In line 2, the placeholder is repaired with recycling of the prefix and the adverb.

- (17) 1 →Y: *dide be-anu agi Sa eh*
 NEG MID-FILL again VOC TAG
 '(you) don't use whatchamacallit again, Sa, right.'
- 2 *be-LKS agi*
 MID-LKS again
 'use LKS⁵(i.e. worksheets) anymore.'
- (BJM01-073-01, 00:12:34–00:12:36, Speaker: Yanti)

The excerpt in (18) is similar, with the demonstrative filler preceded by auxiliaries and prefixed with *be-* 'MID' in line 1. After Hendi initiates repair in line 2, the placeholder is repaired in line 3 with only the *be*-marked predicate.

- (18) 1 →I: *masih nak be:-ini dighi*
 still want MID-FILL self
 'we (lit. self) should have whatchamacallit.'
- 2 H: .. #nak #tuape?⁶
 want what
 'what do (you) want?'
 (0.5)
- 3 I: *be-rakit*
 MID-racket
 'have a racket'
- (BJM01-041-01, 01:07:01–01:07:05, Speakers: Iril, Hendi)

As many studies have discussed, fillers play an important role in word formulation trouble and are inextricably linked to repair (see, e.g., Wouk 2005). Many of the examples of fillers above have demonstrated this connection. Placeholder repair is by nature most often self-initiated, but can be repaired by either party. The previous excerpt in (18), for example, is a case of other-initiated self-repair.

However, this is not the most frequent type of repair. In §4.5, we mention that the vast majority of repairs are instances of same-turn self-repair, as in (11)–(12) and (15)–(17). This may not be surprising given apparent preferences for self-repair in languages like English (Schegloff et al. 1977). There are also instances of

⁵*Lembar Kerja Siswa* 'Student Worksheets' are lessons and activities required by public schools for students.

⁶The # refers to speech that is not clear. Whether Hendi says this phrase or not, it is fairly clear that Iril understood it as an initiation of repair.

other-initiated other-repair, as in (19).⁷ In this excerpt, Leksi and Neti are cooking together. Leksi is explaining what she plans to do with some seasoning when she uses the filler in line 1. Immediately after her use of the filler, Neti initiates and completes repair in line 2, and Leksi acknowledges the repair in line 3.

- (19) 1 →L: *mangke* *aku pacak* (1.0) ***ng-anu-ka=nye[:]***
 consequently 1SG can AV-FILL-CAUS/BEN=3
 'so that I whatchamacallit it.'
- 2 N: *[ng-is]i=nye*
 AV-fill=3
 'fill it.'
- 3 L: *mm*
 mhmm
 'mhmm.'
- (BJM01-158-01, 00:21:13–00:21:16, Speakers: Leksi, Neti)

Fillers also commonly trigger word searches with self-addressed questions, as seen in (14) and (16). A longer word search sequence is provided in (20), where the participants were discussing how expensive buying phone credits was. In the just prior talk, Riska asks how much the taxes are, and Dewi answers in line 1 using the filler. In line 3, Riska agrees with the amount and then initiates repair in line 4 with *pajak* 'taxes'. While Riska initiates repair, Dewi asks a self-addressed question in line 6 overlapping with Riska's repair. Dewi acknowledges the repair in line 7.

- (20) 1 →D: *lapan ribu* *nah [ini=nye baih]*
 eight thousand well FILL=3 just
 'well its whatchamacallit is just eight thousand.'
- 2 A: *[m.]*
 hmm
 'mhmm.'
- 3 R: *[itu=la]:h*
 DEM.DIST=EMPH
 'that's it.'
- 4 *paj-*
 TRUN
 'tax-'

⁷See also the next example in (20).

- (0.5)
- 5 *paj[ak=(ny)e baih kan]*
 tax=3 just right
 'just its taxes, TAG.'
- 6 D: *[tape name=nye]?*
 what name=3
 'what's it called?'
- 7 *au*
 yes
 'yeah.'

(BJM01-081-01, 00:30:20–00:30:26, Speakers: Riska, Dewi, (D)asmi)

Finally, it should be noted that Besemah, as is the case for many insular South-east Asian languages, generally lacks hesitative particles akin to English *uh* or *uhm*. Instead, other cues, such as lengthening and pause, also signal hesitation (see Streeck 1996, Himmelmann 2014). An example of a sound stretch and subsequent pause is found in the example in (21) below. We will see in §4 that these cues are co-present with fillers in a number of instances.

- (21) *ku=kinak-i endik di: (1.2) Lubuk Saung*
 1SG=see-CAUS/LOC N.LI at L. S.
 'I saw the one in (the village of) Lubuk Saung.'

(BJM01-010, 00:05:47–00:05:50, Speaker: Aripin)

In this example, the general locative preposition *di* ‘at’ is lengthened and followed by a fairly long pause before Aripin provides the name of the village *Lubuk Saung*. There are many other examples where pauses and sound stretches accompany fillers. In (14), for example, the demonstrative hesitator *ini* is preceded and followed by pauses, and in (18), the placeholder contains sound stretch.

As mentioned in §1, it is not always possible to determine whether a filler is acting as a hesitator or as a placeholder. Consider the extended example in (22), wherein the demonstrative filler *ini* is used in line 3 and the filler *anu* is used in line 6.

- (22) 1 L: *petang kan aku ke situ*
 evening TAG 1SG to there
 'in the evening, right, I went there.'

- 2 N: *au*
 yes
 ‘yeah.’
- 3 →L: *mane ini kain panjang?*
 where FILL cloth long
 ‘where is uhm/whatchamacallit the sarong?’
- 4 *kain panjang ape kate Mili ni*
 cloth long what word M. DEM.PROX
 ‘what sarong, said Mili.’
- 5 N: *mm*
 mhmm.
 ‘mhmm’
 (1.8)
- 6 →L: *uk- kalu: anu,*
 TRUN maybe FILL
 ‘maybe uhm/whatchamacallit’
- 7 *kain panjang Gina*
 cloth long G.
 ‘Gina’s sarong.’
- 8 *pakai=(ny)e nga Wit*
 pv.use=3 with W.
 ‘Wit wore (it)’
- (BJM01-158-01, 00:55:27–00:55:38, Speakers: Leksi, Neti)

In this excerpt, Leksi is relaying a story about a sarong, *kain panjang* ‘(lit.) long cloth’, to Neti. Before each instance of this word, Leksi uses a filler. In each case, the filler is unaffixed and occurs just before the word she is meaning to say. Under a hesitator analysis, Leksi employs the fillers *ini* and *anu* as delay devices that allow her time to think of the word. Under a placeholder analysis, she employs the placeholder until she selects the more specific word. Without any recycling, evidence from position in the clause, or morphological marking, it is not possible to determine whether these instances are acting as hesitators or as placeholders. In conversations, it is apparent from interlocutors’ responses, however, that this presents no issue, and the filler’s status as a hesitator or a placeholder makes little difference to the progressivity of the interaction (Schegloff 2007). Thus, recycling, position within the clause, and morphological marking are key criteria for identifying placeholders. We return to this issue in §4.3.

Examples like the one in (22) demonstrate that the line that Hayashi & Yoon (2006) draw between placeholder and hesitator is not always so clear. Despite any challenges that this presents to the analyst, however, it does not appear to make any noticeable difference to participants.

3.2 Morphological possibilities of fillers

In the vast majority of cases, fillers in Besemah have the same morphosyntactic properties as either noun roots or intransitive verb roots, both of which may appear as arguments or predicates without any further affixation. Consider the placeholders in (23) and (24), where the placeholder in (23) expresses a nominal argument and the one in (24) a verbal predicate.

In the excerpt in (23), Dewi and Riska are discussing weaving. In line 1, Dewi uses the placeholder *ini*, which occurs as the P argument in an AV construction. It is then repaired with the noun *wi* 'rattan' by both Dewi and Riska in overlapping speech in lines 3 and 4 without any recycling.

- (23) 1 →D: *lemak ng-ambik ini*
pleasant AV-take FILL
'it's nice to take whatchamacallit.'
(0.8)
- 2 R: *au,*
yes
'yeah,'
- 3 D: *[wi.]*
rattan
'rattan.'
- 4 R: *[wi] kapuh ni,*
rattan EXT DEM.PROX
'rattan and the like,'
(BJM01-081-01, 00:20:16–00:20:19, Speakers: Dewi, Riska)

In (24), Neti and Leksi are discussing bananas that they plan to use in their cooking. In line 1, Neti uses the proximal demonstrative as a filler in predicate position after the aspect marker *la* 'PFV'. In line 2, we see recycling of the conditional clause and repair of the placeholder with the predicate *tue* 'old'.

- (24) 1 →N: *ame la ini seghempak die masak*
TOP PFV FILL together 3 ripe
'if (they) are already whatchamacallit, they are both ripe.'

- 2 *ame* (1.0) *la tue nian*
 TOP PFV old very
 ‘if (they) are really old.’
 (BJM01-158-01, 00:52:07–00:52:12, Speaker: Neti)

If a filler is used to express a transitive verb, it appears, with few exceptions, with the causative/applicative suffix *-ka*, which is frequently used to form transitive verbs from nominal or intransitive verbal roots (see §2.1). In (25), the placeholder *anu* is additionally marked by an AV prefix, and in (26), the proximal demonstrative placeholder *ini* in PV is both marked with a proclitic A argument and suffixed with the causative/applicative suffix *-ka*. The small handful of exceptions are discussed below.

- (25) *die galak ng-anu-ka pingging tu=lah*
 3 want AV-FILL-CAUS/BEN butt DEM.DIST=EMPH
 ‘he always whatchamacallited (=push me to the side) (with his) butt.’
 (BJM01-004-01, 00:25:44–00:25:47, Speaker: Dian)
- (26) *galak ng-guring bekayu ni ku=ini-ka*
 want AV-fry cassava DEM.PROX 1SG=PV.DEM.DIST-CAUS/BEN
 ‘(I) like to fry the cassava, (then) I whatchamacallit (=I mix it.)’
 (BJM01-158-01, 00:23:54–00:23:56, Speaker: Leksi)

These properties demonstrate that when *anu* and the demonstrative pronouns are used as placeholders, they are best characterized as falling into one of two word classes, noun or intransitive verb, neither of which requires any additional morphology. Based on the frequencies of morphological patterns, this analysis largely holds in regards to morphological marking; however, there are a few instances wherein *anu* holds the place of a transitive verb without any additional derivational morphology (most commonly the causative/applicative suffix *-ka*) that would be otherwise required for nominal or intransitive verbal roots. As typical of transitive verbs, such a use of the filler can additionally be found with voice morphology *N-* ‘AV’ or *di-* ‘PV’.

The only example in the corpus where the AV prefix occurs with a placeholder not suffixed with *-ka* is shown in line 3 of (27). In this excerpt, Rusmani is telling the other participants about some advice she received to not sell the vegetables she had been growing to an agent, who would then turn around and sell them for a higher price. While the placeholder is not repaired, it is fairly clear from the context that she means selling to an agent.

- (27) 1 R: *anye tape die tu*,
but what 3 DEM.DIST
'but he,'
2 *dide kate=nye*,
NEG say=3
'did not, she said,'
3 → ***ng-anu ke agin***
AV-FILL to agent
'whatchamacallit (=sell them) to an agent,'
(BJM01-071-01, 00:17:00–00:17:04, Speaker: Rusmani)

In (28), Aripin is discussing the process of working with cocoa seeds. In line 2, he uses the placeholder *anu* with the PV prefix *di-* but without the causative/applicative suffix *-ka*. This is only one of a few examples where this occurs. It is noteworthy that the PV prefix only combines with transitive verbal roots; intransitive or nominal roots must be made into transitive verbs with one of the causative/applicative suffixes *-ka* or *-i* in order to combine with the PV prefix.

- (28) 1 A: *tambah di-ghendam*,
add PV-submerge
'even better to submerge (the cocoa seeds).'
2 → ***bukane di-anu***
NEG PV-FILL
'not whatchamacallit (=dry them)'
(BJM01-010-01, 00:26:04–00:26:07, Speaker: Aripin)

These examples demonstrate that as placeholders, *anu* and the demonstrative pronouns have some categorical flexibility, such that they show morphosyntactic properties of transitive verbal roots in addition to intransitive verbal roots and nouns. However, it is important to keep in mind that fillers that show properties of transitive verbs are marginal in the corpus. We should also note that there are instances, such as (12) above, where an unaffixed filler is repaired with a marked transitive verb.

Utilizing the full corpus of Besemah conversations, we see clear patterns with the formal properties of fillers summarized in Table 3. Since *anu* exclusively functions as a filler, we can straightforwardly probe the full corpus for all possible forms and their frequencies. From this, we see that approximately 82% of all forms of *anu* (1,744 tokens) do not occur with any affixation while the remaining

18% (374 tokens) occur with some sort of affixation or cliticization.⁸ Because filler is but one function of demonstrative pronouns, it is only feasible in this study to investigate the morphologically marked forms. As expected, demonstrative pronouns most often occur without any affixation, appearing with some 10,000 instances in the corpus. In the affixed forms of the demonstrative fillers, the proximal demonstrative is far more frequent (137 tokens) compared to the distal demonstrative (21 tokens). When comparing affixed forms of the dedicated filler *anu* and the demonstrative fillers, *anu* with 374 tokens is more than twice as frequent as the demonstratives with 158 tokens combined.

Table 3: Raw frequencies of Besemah fillers occurring with and without additional morphology in the full corpus.

	Demonstratives		
	<i>anu</i>	Proximal <i>ini</i>	Distal <i>itu</i>
Unaffixed	1,744	(<i>unknown</i>)	(<i>unknown</i>)
Affixed	374	137	21
Total	2,118	(<i>unknown</i>)	(<i>unknown</i>)

Table 4 presents a breakdown of these frequencies for the most common forms of filler words in the full corpus. See Table 2 in the previous section for descriptions of the functions of these affixes.

In addition to these more common forms, there are a number of forms that were used only once. For the filler *anu*, this includes a reduplicated form (*anu-anu*), a transitive form with a PV prefix and a locative applicative suffix (*dianui*), a form with the *ke-*-*an* nominalizing circumfix (*keanuan*), and a complex form with the nominalizing circumfix *pe-*-*an* and the numeral prefix *se-* (*sepeanuan*). The AV prefix occurs without a transitivizing suffix only once on a filler (i.e., *nganu*, as seen in (27)). For the proximal demonstrative filler *ini*, the following forms only occur once: *ini* with the reciprocal circumfix *se-*-*an* (*seinian*) and *ini* with the nominalizing circumfix *ke-*-*an* (*keinian*).

⁸Clitics readily occur on both affixed and unaffixed forms of the fillers. In the corpus, the unaffixed filler *anu* occurred with the emphatic clitic *=lah* six times and with the polyfunctional third person pronominal clitic *=nye* 96 times. With the unaffixed proximal demonstrative *ini*, the third person pronominal clitic *=nye* occurred 29 times and the first person pronominal clitic *=ku* four times. These clitics also occur with affixed forms. Fillers with clitics but no prefixes or suffixes are included in the counts for affixed forms in Table 3 but not presented in Table 4.

Table 4: Morphological forms of fillers in the full corpus.

Affix	Demonstrative		
	<i>anu</i>	Proximal	Distal
<i>N-</i> ‘AV’	- <i>ka</i> ‘CAUS/BEN’	131	7
<i>di-</i> ‘PV’	- <i>ka</i> ‘CAUS/BEN’	61	25
	- <i>ka</i> ‘CAUS/BEN’	42	35
<i>di-</i> ‘PV’		6	1
<i>be-</i> ‘MID’		11	20
<i>te-</i> ‘NVOL’		6	5
<i>se-</i> ‘one’		5	1
<i>si-</i> ‘HON’		5	0
<i>-an</i> ‘NMZ’		5	10

These patterns of affixation demonstrate several properties of placeholder fillers, which we revisit in §4. First, transitive forms are almost invariably formed with the benefactive/instrumental applicative suffix *-ka*; the locative/goal applicative *-i* only occurs once in the corpus with *anu* and never with a demonstrative pronoun. This distribution is partially explained by a morphophonological factor that prohibits *-i* from being affixed to *i*-final roots like *ini* ‘DEM.PROX’. This does not, however, explain why it occurs so rarely with *anu*. The very infrequent use of *-i* with placeholders in the corpus is also not explained by its overall frequency. In a sample of four conversations, *-ka* ‘CAUS/BEN’ is overall more frequent with 481 occurrences, but *-i* ‘CAUS/LOC’ is still well represented with 307 occurrences.⁹

Second, as Table 4 shows, affixed forms of *anu* are much more frequent than affixed forms of demonstrative fillers (272 instances of *anu* compared to 104 instances of *ini* and 21 instances of *itu*). However, these differences in frequency are largely due to the disparity between fillers in AV transitive forms with the AV nasal prefix *N-* and applicative suffix *-ka*; this form alone accounts for just under half of all instances of affixed forms of *anu*. It is also striking that with the demonstrative pronoun fillers, the AV form is the least frequent of the transitive verb forms, occurring a total of just eight times as either *nginika* or *ngituka*. To

⁹These numbers are based on counts from the following four conversations, with counts for each conversation in parentheses: BJM01-004 (-*ka* = 106; -*i* = 71), BJM01-008 (-*ka* = 145; -*i* = 41), BJM01-010 (-*ka* = 131; -*i* = 137), BJM01-011 (-*ka* = 99; -*i* = 58).

a lesser extent, PV transitive forms with the prefix *di-* and suffix *-ka* are a little more than twice as frequent with *anu* than the demonstrative fillers.

Finally, other morphological forms of the fillers are far less frequent, making it difficult to interpret their relative frequencies. For example, while the verbal prefix *be-* ‘INTR’ and the nominalizing suffix *-an* ‘NMZ’ are more than twice as frequent with the demonstratives than with *anu*, these numbers are likely too small to extrapolate any sort of conclusion.

Based on a limited typological sample, Podlesskaya (2010: 14) hypothesizes that “if a language has placeholders that can replicate morphology other than nominal, it also has placeholders that replicate nominal morphology, but not vice versa – simply because placeholders are by and large of proNOMINAL origin” (emphasis in the original). The morphological properties of Besemah fillers support this notion, with both verbal and nominal morphology being replicated by placeholders. That nominal morphology is rare in Besemah fillers is simply a reflection of the properties of their source forms, since nominal morphology is largely lacking from Besemah in general. Simply put, verbal morphology is more frequent because affixed verbal fillers are more commonly derived, while nominal uses of the filler do not require the same sorts of derivation.

Another parameter that Podlesskaya (2010) proposes is the extent to which the morphology on fillers replicates the delayed word or constituent. Besemah appears to show mixed results. Filler forms in several cases are fully replicated, as in (29). In this example, Riska is asking Dasmi to help her by giving her child some water. In her request, Riska uses the filler in line 1 with a causative/applicative suffix *-ka*. In line 2, Dasmi asks for clarification, and in line 3, Riska offers a repair whose morphological form exactly replicates that of the filler with the suffix *-ka*.

- (29) 1 →R: *mintak tulung anu-ka dikit yuk*
 request help PV.FILL-CAUS/BEN little.bit older.sister
 ‘please help whatchamacallit, older sister.’
 (0.3)
- 2 D: *luk ma[ne]?*
 how which
 ‘how’s that?’
- 3 R: *[en]juk-ka dikit ngaghi Jep tu*
 PV.give-CAUS/BEN little.bit with J. DEM.DIST
 ‘give (it) to Jep.’
- (BJM01-081-01, 00:26:04–00:26:07, Speakers: Riska, Dasmi)

However, as mentioned above, the causative/applicative suffix *-ka* is all but required to derive transitive verbs with fillers, and there are many examples where the repair does not have the applicative suffix, as in (30) and (31). In fact, repaired transitive verbs lack the causative/applicative suffix much more frequently than have it (although not universally; see (29) above). In (30), Aripin is talking about his cocoa field. In line 1, he uses the filler in a negative imperative construction. The filler is marked by a PV prefix and the causative/applicative suffix *-ka*. In line 2, Damsi repairs the filler, recycling the negative imperative and PV prefix but not the suffix *-ka*.

- (30) 1 →A: *jangan nian di-anu-ka*
 NEG.IMP very PV-FILL-CAUS/BEN
 ‘really don’t whatchamacallit’
 (0.4)
- 2 D: *jangan di-racu[n]*
 NEG.IMP PV-poison
 ‘don’t spray it with poison.’
- 3 A: [a]u
 yes
 ‘yeah.’
- (BJM01-010-01, 00:21:19–00:21:22, Speakers: Aripin, Damsi)

The example in (31) is similar, but the filler and repair are both in AV. In lines 1 and 2, Riska is responding to a question about why the item she is weaving is stiff. She uses a demonstrative filler with the AV prefix, causative/applicative suffix *-ka*, and the P enclitic argument *=nye* referring to the leaves she is using in her weaving. She immediately repairs the filler in line 3 with the AV prefix and P enclitic argument. In this case, the form of the filler and the repair are the same, save the suffix *-ka*.

- (31) 1 R: *emk-*
 TRUN
- 2 → *ng-ini-ka=nye,*
 AV-FILL-CAUS/BEN=3
 ‘(I) whatchamacallit them,’
- 3 *nik me-lipat=e kan,*
 for AV-fold=3 TAG
 ‘folded them, right,’
- (BJM01-081-01, 00:25:17–00:25:20, Speaker: Riska)

3.3 Interim summary

In this section, we have seen that both the dedicated filler *anu* and the demonstrative pronouns are used as either hesitators or as placeholders (§3.1). However, when the filler occurs without any morphology, it is not always possible to categorize its use as a hesitator or placeholder; such cases are thus considered indeterminate. In the corpus of Besemah conversations, fillers most commonly occur without any morphological marking. When they are morphologically marked, they are considered placeholders and may be affixed with a range of prefixes, suffixes, and circumfixes. The most common morphology by far is represented by those affixes that derive transitive verbs, as well as the polyfunctional enclitic =nye that marks third person possession and definiteness among other functions (see §3.2). Finally, we showed that fillers are commonly found in repair sequences. They often co-occur with disfluencies, such as cut-offs and sound stretches. The placeholder and its repair, however, do not always share the same form, especially in the case of transitive verbs. The next section takes a quantitative approach to the study of Besemah fillers to better understand the frequencies of the patterns described in this section.

4 Quantitative analysis of fillers

In an effort to provide a more complete picture of the use of fillers in Besemah interaction, we quantify here the frequencies of many of the features described in §3. We do so by coding each use of the dedicated filler *anu* and the demonstrative pronouns as fillers in a subset of the corpus of Besemah conversations. This section demonstrates several properties of fillers in Besemah. First, while demonstrative pronouns are fairly frequent as fillers, *anu* is more frequent by a ratio of approximately 3:1. Second, fillers are far more frequently used as placeholders than hesitators; this is especially true of demonstrative fillers. Third, disfluencies associated with trouble in word formulation are significantly more common with hesitators than with placeholders. Finally, it is far more frequent for placeholders to be left unrepaired than to be repaired.

4.1 Corpus of Besemah conversations

Thus far, we have considered the full corpus of conversations in Besemah. In this section, we investigate a subset of the full corpus, including six conversations between two or three Besemah speakers each. These conversations were chosen as a fairly representative sample of the larger set of conversations, with an equal

number of men and women ranging in age from 25 to 58 at the time of recording. The subcorpus contains just over 57,000 words in approximately 20,000 intonation units. Short descriptions of these six conversations are presented in Table 5.

Table 5: Conversations included in subcorpus for the quantitative study

ID	Description
BJM01-035	A conversation between two friends and neighbors—Ishan (M, 51) and Darsono (M, 58)—at Darsono’s house.
BJM01-041	A conversation between three friends—Hendi (M, 28), Peter (M, 25), and Hairil (M, 31)—at a friend’s rice paddy.
BJM01-086	A conversation between three friends—Jufri (M, 44), Fikri Tarnizi (M, 36), and Sasli (M, 43)—at Jufri’s rice paddy.
BJM01-116	A conversation between three women—Susianawati (F, 37), Neti (F, 37), and Rusmawati (F, 64)—at Susianawati’s house.
BJM01-125	A conversation between three women—Lismiana (F, 39), Meri (F, 40), and Partiwi (F, 51)—while collecting seeds from carrot flowers.
BJM01-158	A conversation between two friends—Neti (F, 43) and Leksi (F, 35)—while cooking.

Utilizing this subset of conversations, we present an in-depth analysis of the properties of fillers. These six conversations were coded following the schema presented in Appendix A. The properties we coded for fall into the following areas: (i) the type of filler (*anu*, proximal demonstrative pronoun, distal demonstrative pronoun), (ii) the use of the filler (hesitator, placeholder, indeterminate), (iii) its position in intonation units, (iv) the presence of disfluencies with fillers, (v) the morphological marking and syntactic position of placeholders, and (vi) properties related to repair. Qualitative descriptions of these properties were discussed in detail and exemplified in §3.

4.2 Filler forms

Overall, we coded 588 instances of fillers in the six conversations, as summarized in Table 6. The numbers outside of the parentheses represent the frequency count of all forms, while the numbers in parentheses represent the number of fillers that occur with some morphology. This table shows that the filler *anu* is more frequent than the demonstrative pronouns. It also demonstrates that the vast majority of fillers do not occur with any morphology.

Table 6: Frequency counts of Besemah fillers in a subcorpus of six conversations. The total number of fillers found is 588.

Demonstratives		
<i>anu</i>	Proximal <i>ini</i>	Distal <i>itu</i>
439 (71)	120 (28)	29 (3)

If we compare the distribution of fillers in the subcorpus to the distribution in the full corpus presented in §3.2, we see similar patterns in the forms that are available for comparison.¹⁰ For example, in comparing the frequency of affixed and unaffixed forms of *anu*, we found that *anu* occurred without any affixation in approximately 82% of all tokens, while in the subcorpus we see that this number is only slightly higher, at 83% of all instances of *anu* being unaffixed. When comparing the proportion of affixed forms of the demonstrative pronouns and the dedicated filler *anu* in the full corpus to the proportion in the subcorpus, we see that the affixed forms of *anu* make up 70% of all cases in the full corpus and just under 70% of affixed forms in the subcorpus. Thus, based on just these two factors, it looks as if the subcorpus is a good representative of the larger corpus.

4.3 Hesitators versus placeholders

In the subcorpus of six conversations, we coded the primary distinction between *hesitator* and *placeholder* functions for each filler based on the three criteria outlined in §3.1, namely, recycling, position within the clause, and morphological marking. At the same time, we showed in the same section that there are a number of cases where it is not possible to determine whether the filler is being used as a placeholder or a hesitator. These cases were coded as indeterminate. The frequency counts of each are shown in Table 7.

Table 7 demonstrates that while placeholder functions dominate the distribution for all forms, the hesitator use is especially infrequent for proximal demonstrative pronouns, occurring at just under 8% of the proximal demonstrative fillers. For comparison, just under 20% of all instances of *anu* are hesitators. The distal demonstrative pronoun falls somewhere in between, occurring as a hesitator in approximately 14% of instances. However, the relatively infrequent use of the distal demonstrative makes this number somewhat difficult to interpret. One

¹⁰Recall that we were unable to determine (without detailed coding) whether the unaffixed demonstrative pronouns were being used as fillers in the full corpus.

Table 7: Hesitator and placeholder frequency counts of Besemah fillers in the subcorpus.

	Demonstratives		
	<i>anu</i>	Proximal <i>ini</i>	Distal <i>itu</i>
Placeholder	302	99	21
Hesitator	86	9	4
Indeterminate	51	12	4

may argue that these numbers are skewed due to the relatively large number of indeterminate cases. However, even if we considered all indeterminate cases to be hesitators, the same pattern holds: *anu* is overall more frequent and the proportion of hesitators is higher in *anu* than in the demonstrative pronouns.

Several other features are relevant to the distinction between hesitators and placeholders, including the position of the filler within an IU (see e.g., Hennecke & Mihatsch 2022 who find weak correlations between function and prosodic features with fillers in French) and the presence of disfluencies. Hesitators are relatively evenly distributed within IUs across initial (17), medial (20), and final (26) positions. However, there is a clear preference for hesitators to be used in an independent IU (36). Indeterminate fillers are far more frequent in final position (37) than initial (9) or medial (7) positions, with a fewer number occurring in their own IU (12). Placeholder fillers show a different distribution, with clear preference for utterance medial (235) and final (130) positions. They are infrequently encountered initially (46) and only rarely in their own IUs (11). These distributions present a relatively coherent picture that is consistent with differences between placeholders and hesitators. That is, the frequent occurrence of hesitators in independent IUs is consistent with their lack of syntactic integration, and the fact that they are primarily employed as a delay device. The relatively low frequency of placeholders in IU initial position and in independent IUs is consistent with the fact that they are syntactically integrated into their clause.

Disfluencies associated with trouble in word formulation occurred immediately preceding or following the filler in 141 of the 588 occurrences, as shown in Table 8. This table shows that despite hesitators constituting approximately a sixth of filler uses, around a third of the disfluencies coincided with the use of a hesitator (36 of 99) and more than a third with indeterminate cases (25 of 67). This distribution is consistent with analyses that view hesitators primarily as a delay device (Clark & Tree 2002). It is also noteworthy that *anu* is used in all but

three instances of hesitators with disfluencies co-present. This provides further support that there is a clear preference for *anu* to be used as a hesitator.

Table 8: Frequency counts of disfluencies immediately adjacent to the filler by filler type and form.

		Placeholder	Hesitator	Indeterminate
<i>anu</i>	Disfluent	66	33	21
	Fluent	236	53	30
Proximal <i>ini</i>	Disfluent	10	1	-
	Fluent	89	8	12
Distal <i>itu</i>	Disfluent	4	2	4
	Fluent	17	2	-

The most common disfluencies were pauses (32), truncation (33), and lengthening (see (18) above) preceding or following the use of the filler. In a number of cases, these disfluences co-occur in the same utterance. Pauses were the most common, occurring 64 times, followed by truncation with 55 occurrences and lengthening with 29 occurrences in the subcorpus.

- (32) *la beghape kali die ni .. anu .. ny-(s)ayur ni*
 already many time 3 this FILL AV-vegetable this
 'he already multiple times, umm, planted these vegetables.'
 (BJM01-086, 00:02:47–00:02:50, Speaker: Sasli)

- (33) 1 →D: *ka- katah alap=e: cu- [anu=nye]:*
 TRUN how good=3 TRUN FILL=3
 'how beautiful were her whatchamacallit.'
- 2 I: *[au]*
 yes
 'yeah.'
- 3 D: .. *tumat=e*
 tomato=3
 'her tomatoes.'
- (BJM01-035-01, 00:28:22–01:07:05, Speakers: Darsono, Ishan)

4.4 Placeholder syntactic properties

The examples in §3 demonstrate how placeholders in Besemah can be predicates, as in (11) and (12) above, or arguments, as in (15) and (16) above. It is also possible for placeholders to be the object in a PP, as in (34), or a modifier in an NP, as in (35).

- (34) 1 D: *jeme* (H) *mane*,
person which
'which person,'
2 → *jalan ke anu tu eh*
walk to which DEM.DIST TAG
'walked to whatchamacallit, right.'
(BJM01-035-01, 01:07:00–01:07:03, Speaker: Darsono)

- (35) 1 →U: *mesin .. ini*,
machine DEM.PROX
'whatchamacallit machine,'
2 J: .. *au*,
yes
'yeah,'
3 U: *mesin padi kecik eh*
machine rice small TAG
'the small rice (milling) machine.'
(BJM01-086-01, 00:35:44–00:35:47, Speakers: Ujang, Jupri)

In (34), Darsono is launching into a story but having trouble locating the name of the person (line 1) and where this person is going (line 2). In line 2, the placeholder *anu* occurs as the object of the preposition, occurring between the preposition *ke* 'to' and the demonstrative determiner *tu*. It is eventually repaired by Darsono after a lengthy word search. In line 1 of (35), Ujang is discussing rice milling and employs the noun *mesin* 'machine', which is followed by the proximal demonstrative placeholder *ini*. In line 3, the head noun is recycled and the placeholder in modifier position is replaced. The resulting NP is *mesin padi kecik* 'small rice (milling) machine'.

The frequencies in which placeholders occupy these positions in the clause are shown in Table 9. The numbers within the parentheses represent the morphologically marked forms, while the numbers outside of the parentheses represent the total number of forms irrespective of morphological marking.

Table 9: Placeholder frequency counts of Besemah fillers by syntactic position.

	Argument	Modifier	PP-Object	Predicate
<i>anu</i>	80 (26)	25 (1)	19 (0)	178 (43)
Proximal <i>ini</i>	17 (4)	6 (1)	14 (1)	62 (22)
Distal <i>itu</i>	6 (1)	2 (0)	4 (1)	9 (1)
Total (422)	103	33	37	249

Table 9 demonstrates how Besemah placeholders have a wide syntactic distribution, which may be unsurprisingly similar to colloquial Indonesian as described in Wouk (2005) (see also Podlesskaya 2010: 13–14). The syntactic positions we have coded in Table 9 largely overlap with traditional word classes. Where placeholders in our coding are repaired, arguments and objects of prepositions are exclusively nouns or NPs while predicates are almost exclusively verbs, with a single exception where the predicate in a cleft construction is repaired by a noun. Modifiers are repaired primarily by nouns or NPs (e.g. possessors), with a small number of instances repaired by verbs.¹¹

While it is possible to identify these positions without morphological marking, certain affixes and clitics, if present, allow for rather transparent interpretation of the placeholder’s syntactic status. As described in §3.2, strictly nominal or verbal morphology most clearly provides evidence for this. There are only two enclitics that occur with both nominal and verbal forms: the focus enclitic =*lah* and the enclitic =*nye*, which can both attach to nominal arguments (marking third person possession or definiteness) and to verbal predicates (referencing the A argument in PV or the P argument in AV). In the latter case, the filler is always occupying the position of a transitive verb and is thus suffixed with the causative/applicative -*ka* ‘CAUS/BEN’, so there is no confusion when coding these examples.

What is perhaps most striking about the syntactic positions of placeholders in Besemah is the skewing towards predicates in all forms; predicates make up roughly 60% of all placeholders. Prototypically nominal positions, such as arguments and objects in PPs, make up the majority of the remaining positions of placeholders followed by nominal modifiers. While Podlesskaya (2010) proposes that “[i]t is more common for placeholders to substitute nominal constituents”

¹¹Besemah, like many Austronesian languages, does not distinguish adjectives from verbs (see Mosel 2023).

across languages, it is interesting that in Besemah the most frequent positions are associated with verbal elements.

Comparisons among the different placeholder forms do not reveal any significant differences in their morphosyntactic properties. Predicates occur with similar frequencies across the fillers, although with only 21 total occurrences, it is difficult to draw any definitive conclusions about distal demonstratives. The dedicated filler *anu* occurs more frequently in argument and modifier positions, while the proximal demonstrative *ini* occurs more often as the object of a preposition. These differences are not necessarily insignificant, but it is difficult to draw any strong conclusions based upon the distributions.

As expected, verbal morphology only occurred with placeholders in predicate positions, and of the placeholders in predicate position, a quarter of the instances of the dedicated filler *anu* (43 of the total 178) occurred with verbal morphology, while a third of the proximal demonstratives *ini* (22 of the total 62) occurred with verbal morphology. Distal demonstrative placeholders occur only once with explicit verbal morphology in the corpus. As was the case in the full corpus (§4.1), transitive verbal morphology was the most commonly employed among all fillers. The most frequent form is the dedicated filler *anu* with both the AV prefix *N-* and causative/applicative suffix *-ka*, *ng-anu-ka*.

Nominal morphology primarily included the third person possessive enclitic, and for demonstrative placeholders, this was the only morphological marking. For *anu*, there were a few additional forms, but third person enclitics dominated the distribution.

4.5 Repair properties

Table 10 compares the relative frequencies of placeholders that are repaired with those that are not. Most striking is the fact that all forms are more likely to be left unrepairs than repaired. That is, repair only occurs in one-third of all instances of placeholders. The dedicated filler *anu* is the least likely to be repaired. This is perhaps unsurprising, though, since, as opposed to the demonstratives, *anu* is solely used as a filler. The demonstrative pronouns in their placeholder functions are more likely to be repaired, but are still outnumbered by instances where the placeholder is left unrepairs.

There is also a somewhat expected interaction between repair and disfluency: when disfluency is present, repair is more likely to occur. Table 11 shows that placeholders from *anu* are almost equally likely to be repaired as not when disfluency is present, compared to repair occurring just one-fourth of the time when

Table 10: Placeholder frequency counts of Besemah fillers by repair.

	Repaired	Unrepaired
<i>anu</i>	87	215
Proximal <i>ini</i>	35	64
Distal <i>itu</i>	9	12
Total (422)	131	291

there are no disfluencies. This pattern also appears to hold for demonstrative pronouns, although their use with disfluencies is relatively infrequent.

Table 11: Frequency counts of disfluencies with repaired and unrepaired placeholders in the subcorpus of Besemah.

	Disfluent		Fluent	
	Repaired	Unrepaired	Repaired	Unrepaired
<i>anu</i>	32	36	57	178
Proximal <i>ini</i>	6	4	29	60
Distal <i>itu</i>	3	1	6	11

4.6 Differences between *anu* and the demonstrative pronoun fillers

From the discussion above, there are few distinguishing patterns between the filler uses of *anu* versus the demonstrative pronouns, aside from the overall higher frequency of *anu*. One significant difference is that while all three forms are more frequently used as placeholders rather than hesitators, *anu* is much more likely to be used as a hesitator than the demonstrative pronouns are. However, an important factor that has yet to be addressed is the individual preferences of speakers. The variation in these preferences is apparent from the frequencies of different forms in each of the conversations, as displayed in Table 12.

This table demonstrates that in all but two conversations, the use of *anu* dominates the distribution. In the two sessions where demonstrative pronouns are more frequent (BJM01-041-01, BJM01-158-01), *anu* is still relatively frequent.

Looking at the same frequencies among the 17 speakers in these six conversations reveals that a combination of age and time spent outside of the Besemah

Table 12: Frequency counts of different forms by conversation. Conversations that are shaded favor demonstrative pronouns as fillers.

	<i>anu</i>	Proximal <i>ini</i>	Distal <i>itu</i>
BJM01-035-01	135	–	14
BJM01-041-01	29	39	2
BJM01-086-01	34	7	1
BJM01-116-01	103	2	2
BJM01-125-01	102	12	3
BJM01-158-01	36	60	7

region appear to correlate with an increased use of demonstrative pronouns. Consider the frequency counts of filler use for each speaker in the subcorpus alongside some basic demographic factors as shown in Table 13.¹² Younger speakers of the sample (born after 1975, the median birth year) tend to use more demonstrative pronouns than older speakers, especially if they spent significant time outside of the region. For older speakers (born before 1975), on the other hand, the use of *anu* is far more frequent than the use of demonstrative pronouns. There are several apparent exceptions to these patterns.

Two interesting participants, Sasli and Peter, demonstrate the importance of both age and time outside the region. While Sasli was born before the median birth year, he lived outside of the region for six years, and still he exclusively uses *anu*. Peter, despite being the youngest speaker of this sample, had not lived outside the region and also exclusively uses *anu*. This supports the hypothesis that it is the combination of age and time outside of the region that correlates with the use of demonstrative pronouns.

In the cases of Meri and Ujang, we see exceptional patterns. That is, Meri spent significant time outside of the region and her birth year is the same as the median. However, she very infrequently uses the demonstrative pronouns as fillers. This pattern is especially stark when compared with Neti, who was born in 1975. She spent roughly the same time outside of the region, and has the second lowest percentage of *anu* use. Ujang shows the inverse of Meri, preferring to use demonstrative pronouns as fillers. He was born after the median year, but

¹² Although the principal number of participants in the subcorpus totaled 17, there were a total of 20 speakers who appeared in the subcorpus. The additional participants were family members or friends who momentarily joined the conversation. Two participants who fall into this category used a total of six fillers. In an effort not to skew the results with these small numbers, these participants are not included in Table 13.

Table 13: Frequency counts of the use of different fillers with rows sorted from the lowest proportion of *anu* to the highest alongside speaker's name, year of birth, gender, and time outside of the Besemah region in years.

Birth year	Name	Gender	Outside region	<i>anu</i>	<i>ini</i>	<i>itu</i>	% of <i>anu</i>	
1981	Hendi	m		5	1	20	1	5%
1975	Neti	f		4	3	43	2	6%
1979	Ujang	m		0	2	6	1	22%
1983	Iril	m		12	7	19	1	26%
1983	Leksi	f		4	33	17	5	60%
1964	Partiwi	f		0	34	7	2	79%
1963	Ishan	m		0	50	—	12	81%
1975	Lis	f		0	30	3	—	91%
1975	Meri	f		4	38	2	1	93%
1971	Jupri	m		1	14	1	—	93%
1978	Susianawati	f		0	45	1	2	94%
1951	Rusmawati	f		0	37	1	—	97%
1956	Darsono	m		0	83	—	1	99%
1978	Neti	f		0	21	—	—	100%
1984	Peter	m		0	18	—	—	100%
1972	Sasli	m		6	18	—	—	100%

has not spent significant time outside of the region. However, with less than 10 instances of fillers, it is difficult to draw any reliable conclusions from his use of fillers.

The overall picture that these distributions produce is that the majority of speakers have a bias for the dedicated filler *anu*, with a minority showing a preference for the demonstrative pronouns. Those that favored the demonstrative pronouns were born in 1975 or later with a mix of men and women. Those that favored *anu* comprised a more diverse group with younger and older speakers. One possible explanation for these patterns is language contact. While the varieties of Malay/Indonesian with which Besemah is in contact have not yet been described, reports from colloquial varieties of Jakarta Indonesian show that demonstrative pronouns are most commonly employed as fillers (Wouk 2005, Williams 2010). In fact, both Wouk (2005) and Williams (2010) do not report any use of *anu* in their data, even though it has been reconstructed for Proto Malayic (Adelaar 1992).

5 Conclusion

Studies of fillers, such as Enfield (2003) and Hayashi & Yoon (2006), have rightly pointed out how important fillers are for interaction, serving a wide range of purposes for participants, none more important than dealing with trouble in word formulation. At the same time, interactional data are essential for understanding how fillers are employed by participants. It is especially important to investigate fillers in a range of typologically diverse languages. With few studies of Austronesian languages (with notable exceptions, such as Nagaya 2022, Tanangkingsing 2022), this chapter shows how fillers are used by speakers of Besemah in a relatively large corpus of conversations. In doing so, we have shown how several of the features that Hayashi & Yoon (2006) argue are defining features of fillers cannot be characterized as such in Besemah. For Hayashi & Yoon (2006), the distinction between placeholder and hesitator fillers appears to be clear-cut. Placeholders are referential and syntactically integrated, and they are “replaced” by a more specific word. Hesitators, on the other hand, are non-referential and not syntactically integrated, and the ‘outcome’ following the use of a hesitator (unlike repair) has little correspondence to it.

It may be that Fox’s (2010: 2) characterization that placeholders “fulfill the syntactic projection of the turn so far, rather than simply delaying the next word due,” better captures the distinction. We showed that ‘fulfilling a syntactic projection’ and ‘delaying the next word due’ are not necessarily clear in languages like Besemah with looser syntactic structures. Furthermore, such a distinction may not be consequential for speakers to carry out their interactions. Despite this, there are many instances where the use of hesitators is evident; they do not fulfill a syntactic projection and appear to simply delay the next word due. Based on this evidence, we would advocate for a position that recognizes prototypical placeholders that are morphologically marked and syntactically positioned to fulfill a projection. As fillers move farther from this prototype, losing morphological marking, occurring in less articulated syntactic positions, and being left unrepaired, they become less prototypical. Cases that we marked as indeterminate are positioned in this gray area between fillers that are clearly placeholders and those that do not fulfill a projection.

We have also shown that even though Besemah has a dedicated filler *anu*, many speakers employ demonstrative pronouns as fillers with the same morphosyntactic properties and much of the same distributional properties. However, aside from the fact they are overall less frequent than *anu*, there are intriguing differences in their frequencies along several dimensions. For example, (i) demonstrative pronouns rarely occur in AV, although this is common for *anu*,

(ii) they are less likely to occur as hesitators than placeholders, (iii) they are more likely to be repaired than *anu* placeholders. While these properties are certainly interesting, the use of demonstrative pronoun fillers may be best explained by contact with other varieties of Malay/Indonesian, as evidenced by the preferences of younger speakers who have lived for some time outside of the Besemah region. In other words, few Besemah speakers utilize a roughly equal proportion of the different filler forms; they tend to strongly prefer *anu* or the demonstrative pronoun *ini*.

Abbreviations

1	first person	N.LI	Light Noun
2	second person	LOC	Locative Applicative
3	third person	MID	Middle
AV	A-Voice	NEG	Negative
BEN	Benefactive Applicative	NMZ	Nominalizer
CAUS	Causative	NVOL	Non-volitional
DEM	Demonstrative	PFV	Perfective
DIST	Distal	PL	Plural
DISTR	Distributive	PROX	Proximal
EMPH	Emphatic	PV	P-Voice
EXT	General Extender	RDP	Reduplication
FILL	Filler	REC.PST	Recent past
HON	Honorific	SG	Singular
IMP	Imperative	TAG	Tag question
INTS	Intensifier	TOP	Topic
INT	Interrogative	TRUN	Truncation
		VOC	Vocative

Acknowledgements

We would like to thank two anonymous reviewers and the editors for valuable feedback on this chapter. We would also like to acknowledge Hendi Feriza who has provided his expertise in the transcription, translation, and analysis of the Besemah conversations for the purposes of this paper. We would like to thank our research counterpart in Indonesia, Yanti (Atma Jaya Catholic University of Indonesia), and to the National Research and Innovation Agency (BRIN) for allowing us to carry out this research. This research is based upon work supported by

the National Science Foundation under Grant BCS-1911641. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The corpus of Besemah was compiled by the first author with funding support from a Fulbright scholarship, Blakemore Freeman Fellowship Language Grant, and Fulbright-Hays Doctoral Dissertation Research Abroad scholarship with permission to conduct research on Besemah from the Ministry of Research and Technology in Indonesia. We are especially grateful to all of the Besemah people who have participated in the conversations in the corpus and supported this research more generally, most especially, Asfan Fikri Sanaf, Ken-cana Dewi, Sarkani, and Hendi Feriza.

References

- Adelaar, K. Alexander. 1992. *Proto-Malayic: The reconstruction of its phonology and parts of its lexicon and morphology* (Pacific Linguistics Series C - 119). Canberra: The Australian National University.
- Adelaar, K. Alexander. 2005. Structural diversity in the Malayic subgroup. In K. Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar* (Routledge Family Language Series), 202–226. New York: Routledge.
- Billings, Blaine & Bradley McDonnell. 2025. Fillers in the gray: Indeterminacy in their use in everyday conversation. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 211–250. Berlin: Language Science Press. DOI: ??.
- Blust, Robert. 2013. *The Austronesian languages* (Pacific Linguistics Series C 602). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Blust, Robert, Stephen Trussel & Alexander D. Smith. 2023. *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2) [dataset]*. DOI: 10.5281/zenodo.7741197.
- Chen, Victoria & Bradley McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5(1). 173–195. DOI: 10.1146/annurev-linguistics-011718-011731.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.

- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane Anne Edwards & Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale: Lawrence Erlbaum Associates, Publishers.
- Enfield, N. J. 2003. The definition of *what-d’you-call-it*: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Ewing, Michael C. 2019. The predicate as a locus of grammar and interaction in colloquial Indonesian. *Studies in Language* 43(2). 402–443. DOI: 10.1075/sl.17062.ewi.
- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders*, 1–10. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.01fox.
- Hayashi, Makoto. 2003. *Joint utterance construction in Japanese conversation* (Studies in Discourse and Grammar 12). Amsterdam: John Benjamins. DOI: 10.1075/sidag.12.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Hennecke, Inga & Wiltrud Mihatsch. 2022. Between placeholder and filler: Degrees of prosodic prominence of French *machin* and *truc*. *Pragmatics & Cognition* 29(2). 297–323. DOI: 10.1075/pc.22003.hen.
- Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In K. Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar* (Routledge Family Language Series), 110–181. New York: Routledge.
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960. DOI: 10.1353/lan.2014.0105.
- McDonnell, Bradley. 2008. *Besemah (BJM01)*. Collection BJM01 at catalog.paradisec.org.au [Open Access]. DOI: 10.4225/72/56E823DF6C2D4.
- McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. University of California, Santa Barbara. (Doctoral dissertation).
- McDonnell, Bradley. 2022. Universal quantifiers, focus, and grammatical relations in Besemah. *Studies in Language* 47. 422–462. DOI: 10.1075/sl.20060.mcd.

- McDonnell, Bradley & Christina L. Truong. 2024. Applicative constructions in languages of Western Indonesia. In Fernando Zúñiga & Denis Creissels (eds.), *Applicative constructions in the world's languages*, 971–1004. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110730951-028.
- McDonnell, Bradley, Jiang Wu, Timothy McKinnon & Alexander Adelaar. 2024. The Malayic languages. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*, 431–454. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198807353.003.0029.
- Mosel, Ulrike. 2023. Word classes in Austronesian languages. In Eva van Lier (ed.), *The Oxford handbook of word classes* (Oxford Handbooks in Linguistics), 771–787. New York: Oxford University Press.
- Nagaya, Naonori. 2022. Beyond questions: Non-interrogative uses of *ano* ‘what’ in Tagalog. *Journal of Pragmatics* 190. 91–109. DOI: 10.1016/j.pragma.2022.01.007.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages* (Pacific Linguistics 646). Berlin: De Gruyter Mouton.
- Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in Conversation Analysis I*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511791208.
- Schegloff, Emanuel A., Gail Jefferson & Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53(2). 361–382. DOI: 10.2307/413107.
- Streeck, Jürgen. 1996. A little Ilokano grammar as it appears in interaction. *Journal of Pragmatics* 26. 189–213.
- Tanangkingsing, Michael. 2022. Pragmatic functions of versatile *unsa* ‘what’ in Cebuano: From interrogative pronoun to placeholder to stance marker. *Journal of Pragmatics* 193. 59–75.
- Truong, Christina L. & Bradley McDonnell. 2022. Neglected functions of Western Indonesian applicatives. In Sara Pacchiarotti & Fernando Zúñiga (eds.), *Applicative morphology: Neglected syntactic and non-syntactic functions*, 405–436. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110777949.
- van Valin, Robert D. Jr & Randy J. LaPolla. 1997. *Syntax, structure, meaning and function*. Cambridge: Cambridge University Press.

- Williams, Nicholas. 2010. Toward a linguistic anthropological account of deixis in interaction: *ini* and *itu* in Indonesian conversation. *Colorado Research in Linguistics* 22. DOI: 10.25810/0GHQ-WK77.
- Wouk, Fay. 2005. The syntax of repair in Indonesian. *Discourse Studies* 7(2). 237–258.

Appendix A Coding schema for quantitative study

ID	Description	Levels
Form	What is the full morphological form of the filler?	e.g., <i>ini=nye</i>
Root	What is the root of the filler?	<i>anu, ini, itu</i>
Function	What is the function of the filler?	placeholder, hesitator, indeterminate
IU position	What position within the Intonation Unit (IU) does the filler occupy?	initial, medial, final, sole
Syntactic position	What position in the clause does the filler occupy?	predicate, argument, modifier, PP object, adjunct
Disfluency	Are there any disfluencies immediately preceding or following the filler?	truncation, sound stretch, pause, none
Repair	Is the placeholder repaired?	yes, no, NA
Recycling	Is any part of the phrase recycled?	yes, no, NA
Repair location	How many turns after the placeholder is used is it repaired?	0-5, NA
Repair location	How many IUs after the placeholder is used is it repaired?	0-9, NA
Repair initiator	Who initiates repair of the placeholder?	self, other, NA
Repairer	Who repairs the placeholder?	self, other, NA

Chapter 6

Fillers in the gray: Indeterminacy in their use in everyday conversation

Blaine Billings^a & Bradley McDonnell^a

^aUniversity of Hawai'i at Mānoa

Previous studies that focus on the use of demonstrative and interrogative pronouns as fillers have identified two primary functions, a placeholder function and an (interjective) hesitator function. At the same time, these studies have demonstrated both that placeholders and hesitators can often be clearly distinguished and that the majority of placeholders are eventually repaired. In this chapter, we describe how demonstrative and interrogative pronouns are used as fillers in everyday conversations in Nasal, an Austronesian language of Western Indonesia. Instead of clear-cut boundaries between these functions, we show (i) how distinctions between placeholder and hesitator functions are often indeterminate and (ii) how many instances of the placeholder function do not result in repair. Despite this indeterminacy and lack of placeholder repair, there are no apparent issues for participants in the course of their interaction.

keywords: Austronesian, fillers, conversation, repair, word-formulation trouble

1 Introduction

Demonstrative and interrogative pronouns are among the most common – if not *the* most common – source for fillers cross-linguistically (Podlesskaya 2010), and the uses of different types of pronouns in these roles appear to be well motivated. In dealing with trouble in word-formulation, speakers commonly employ them as fillers, which according to Hayashi & Yoon (2006), fall into two categories: placeholders and (interjective) hesitators. Placeholders, as the name suggests, take the place of an unspecified constituent, be it an individual lexical item or



phrase, which is typically described as being subsequently “replaced” by a more specific lexical item or phrase. This characterization leads Podlesskaya (2010: 11) to refer to placeholders as performing a “preparatory substitute” function. Placeholders are thus syntactically integrated and often carry some or all of the requisite morphology (Hayashi & Yoon 2006: 489). Hesitators, on the other hand, are used in sequences of word-formulation trouble to delay “the production of the next item due in the course of the unfolding utterance” (Hayashi & Yoon 2006: 507). They are not syntactically integrated and thus not repaired by a more specific lexical item.

An important factor that motivates the use of demonstrative pronouns as fillers is their focusing/pointing role, which draws the speaker’s and hearer’s focus on the same, yet unspecified referent. This property helps to explain how demonstrative pronouns are commonly taken up as a resource for dealing with word-formulation trouble because, as Hayashi & Yoon (2006: 515) put it, “a placeholder demonstrative creates a prospective link to a subsequent specification of the referent and focuses the hearer’s attention on it.” For hesitators, Hayashi & Yoon (2006) propose that demonstratives lose referential features via a process of pragmaticization, developing not only into hesitators but also discourse markers with a wide variety of functions (see Hayashi & Yoon 2006: 525–529 for discussion of discourse functions of demonstratives in Mandarin, Japanese, and Korean as well as Næss et al. 2020 for a broader discussion of discourse functions of demonstratives).

Interrogative pronouns have received less attention, but in a description of placeholders derived from interrogative pronouns in two Caucasian languages, Udi and Agul, Ganenkov et al. (2010) point out close similarities between the use and form of self-addressed questions on the one hand and placeholders on the other. In fact, they state that there are cases where distinguishing the two is impossible. Such contexts help explain how interrogative pronouns may have developed non-interrogative functions as placeholders. The development of hesitators from interrogative pronouns could operate through a similar process of pragmaticization as with demonstrative pronouns, although it is not difficult to imagine that interrogative pronouns follow a distinct pathway of change where self-addressed questions need not become a placeholder before developing hesitator functions.

Although placeholder and hesitator are treated as two distinct functions in the descriptions of Mandarin, Japanese, and Korean by Hayashi & Yoon (2006), in practice it is unclear to what extent these functions can be distinguished in the world’s languages. In Nasal, for example, the lack of obligatory morphological

marking and frequent lack of repair for placeholders renders two of the otherwise more reliable criteria for delineating these two functions useless (see also McDonnell & Billings 2025 [this volume] for a similar issue in Besemah). Indeterminacy is even more pronounced with apparent hesitators, where the line between a hesitator and a self-addressed question used as a delay device is often blurry (see Ganenkov et al. 2010: 33 for a similar indeterminacy in the use of the interrogative pronoun as a placeholder and in self-addressed questions). In a broader respect, there is a third level of indeterminacy in analyzing whether a demonstrative or interrogative pronoun is used in a filler function at all. When morphological marking does not appear on one of these function words, it is often difficult to determine whether the function word is serving one of its more typical deictic functions (not filler), taking the place of some yet unspecified referent (placeholder), or simply delaying speech (hesitator).

Despite their presence in a wide variety of typologically diverse languages, there are relatively few studies on the use of demonstrative and interrogative pronouns as fillers in everyday conversations (see Hayashi & Yoon 2010: 58). This is particularly true in Austronesian languages where there are few studies dedicated to the use of fillers (Nagaya 2022 for Tagalog, Dimock 2010 for Navahaq, and Williams 2010 for Indonesian being notable exceptions). There are several studies that describe how Austronesian languages use demonstratives in word-formulation trouble. For instance, Wouk (2005), in discussing repair in Indonesian conversation, has shown that demonstrative placeholders play a significant role in same-turn self-repair. Ewing (2005), in discussing demonstrative pronouns in Cirebon Javenese conversation, describes how they are used as hesitation markers. In a grammar of Papuan Malay, Kluge (2017) briefly describes how demonstrative pronouns are utilized as both placeholders and hesitators.

Across the Austronesian family, however, interrogative pronouns are more commonly described as a source for fillers and have even been reconstructed as fillers for higher-level subgroups, such as Proto Malayo-Polynesian *a-nu ‘thing whose name is unknown, avoided, or cannot be remembered; what?’ (Blust 2013: 516, Blust et al. 2023). In quite a few languages, this filler has lost its interrogative function, and the form has become a dedicated filler word (see McDonnell & Billings 2025 [this volume]). Within this context, Nasal (glottocode: nasa1239), an Austronesian language of Indonesia, presents a particularly interesting case wherein both the demonstrative pronouns and the interrogative pronoun are used in placeholder and hesitator functions. Unlike well-known Austronesian languages like Indonesian that use demonstrative pronouns as fillers, Nasal has a rather complex three-way demonstrative system, and the filler functions of the interrogative pronoun *api* ‘what’ have not fully lexicalized. Instead, it shows

how *api* ‘what’ still occurs in word-formulation trouble both by itself as a lexical hesitator as well as in phrasal expressions of self-questioning in word searches. Potentially more surprising is that, unlike the examples examined in Hayashi & Yoon (2006), many instances of fillers in Nasal conversation are indeterminate in terms of their placeholder or hesitator function. That is, in a number of instances the function of the filler cannot be straightforwardly analyzed, either because of lack of morphological marking, lack of repair, or homophony with other forms. While the indeterminacy between these two functions does not present any issues for the participants in the course of their interaction, it raises a number of issues in the analysis and calls into question whether the divide between these functions is as sharp as Hayashi & Yoon (2006) describe.

This chapter is organized as follows. The following section, §2, provides a brief introduction to Nasal, its morphosyntax, and the corpus used for this study. §3 gives an overview of those lexical items that operate as fillers in Nasal and their various uses. §4 and §5 describe hesitator and placeholder uses of these fillers, respectively. §6 discusses the distribution of these fillers in the larger corpus. Finally, §7 and §8 provides some broader implications of our analysis and concluding remarks.

2 Nasal

Nasal is an underdocumented language of the Sumatran subgroup within the Malayo-Polynesian branch of the Austronesian language family and is spoken by approximately 3,000 people along the southwestern coast of Sumatra (Billings & McDonnell 2024). Documentation is limited to a few wordlists and texts (Stokhof 1987, Kasim et al. 1987, Anderbeck & Aprilani 2013) along with the authors’ ongoing documentation project (McDonnell 2017, McDonnell et al. 2019). Nasal is surrounded by much larger speaker populations of neighboring Malayic and Lampungic languages. While the languages of these three subgroups of Malayo-Polynesian—Sumatran, Malayic, and Lampungic—share a number of the morphosyntactic properties found in the languages of the western Indonesia region, they also show considerable variation from one another in nearly all domains, from phonology to lexicon to syntax (see McDonnell et al. 2024, McDonnell & Truong 2024b). At the same time, neighboring Lampungic and Malayic languages have considerably affected Nasal, likely as the result of widespread multilingualism (McDonnell to appear). This has led to the development of a grammar that is strikingly similar to both neighboring Lampungic and Malayic languages (i.e. South Barisan Malay, Kaur) with a high number of loan words from both subgroups. For the discussion here, this is perhaps most relevant in the domain of

demonstratives and interrogatives, where the Nasal system stands in stark contrast to Malayic but has clear affinities with Lampungic. Although there remains a distinctively Nasal core, these multiple layers of borrowing complicate the picture of Nasal.

In the remainder of this section, we provide a brief overview of the basic structural properties of Nasal as a foundation for the study of Nasal fillers and introduce the corpus used for this study.

2.1 Basic morphosyntactic properties

In Nasal, as in other languages of Western Indonesia, predicates commonly occur without explicit arguments, and while the absence of arguments has traditionally been analyzed as a form of ellipsis, recent studies that take an interactional approach to these languages do not assume that arguments are elided (e.g. Ewing 2019). Instead, these approaches show how such minimal structures are the norm, as in (1), where the verbal predicate in line 3 occurs without any arguments.¹

- (1) 1 A: *nyak so hago m-inum*
1SG this want INTR-drink
'I want to drink (something)'
- 2 U: *akuk=do*
take=EMPH
'take (it)'.
- 3 A: *mak pandai ng-akuk,*
NEG able AV-take
'(I) can't get (anything)',
(BJM02-002, 01:07:53–01:07:56, Speakers: Arma, Upik)

As in many other Austronesian languages (see Mosel 2023), word classes can be difficult to distinguish in Nasal where a number of roots may be considered flexible or unspecified for word class. The word class of lexical roots is thus entirely determined by their syntactic role in a given utterance. While nominal and verbal morphology draws a clear word class divide in complex forms, affixes

¹We use a simplified and updated version of Discourse Transcription (Du Bois et al. 1993). Each new line represents an Intonation Unit (IU) with end marks on each representing different contours: ‘.’ is a ‘final’ contour, ‘,’ is a ‘continuing’ contour, and ‘?’ is an ‘appeal’ contour. Brackets represent overlapping speech. An en dash ‘–’ represents a truncated word, an em dash ‘—’ represents a truncated IU, ‘.’ represents lengthening, ‘..’ represents a short pause, (TSK) represents a click of the tongue, and (%) represents glottalization.

may operate on roots from either class (i.e. verbal affixes may combine with both nominal and verbal roots and nominal affixes may combine with verbal and nominal roots). In the examples, therefore, verbal roots may appear with a nominal gloss and vice-versa.

Nouns take no grammatical marking for gender, number, or case, and noun phrases minimally consists of a noun. Numerals and numeral classifiers typically precede the noun, with other modifiers, like demonstratives, following the noun. Possessors follow the head noun, and possession can also be indicated by pronominal enclitics *=ku* ‘1SG’, *=mu* ‘2SG’, and *=nyo* ‘3SG’. As is common for languages of the region, the third-person singular enclitic *=nyo* additionally expresses definiteness among a number of other functions (see, for example, Gil’s 2003 analysis of Riau Indonesian *=nya*). Relative clauses follow the head noun and are marked by the relativizer *sai*; they are commonly headless.

Intransitive verbs are often unmarked (2) but may take one of several prefixes, such as the non-volitional prefix *te-* ‘NVOL’ or the middle prefix *be-* ‘MID’ (3). These prefixes derive verbal predicates from nominal or verbal roots.

- (2) *yo khatung dijo jo*,
3SG come here earlier
'he came here before,'

(BJM02-007, 00:48:44–00:48:45, Speaker: Eka)

- (3) *sawah kito be-wayil*
field 1PL.INCL MID-water
'our field has water'

(BJM02-005, 00:34:28–00:34:30, Speaker: Mas)

Transitive verbs in Nasal make use of a two-way voice distinction between an A-Voice (AV), marked by the so-called nasal prefix *N-* ‘AV’ (4), and a P-Voice (PV), marked by the prefix *di-* ‘PV’ or a bare verbal stem (5).² Similar to intransitive prefixes, these prefixes derive verbal predicates from both nominal and verbal roots. In both voice constructions, the Primary Argument (A in AV, P in PV) has privileged access to certain syntactic operations (e.g. relativization), and the Secondary Argument (A in PV, P in AV), depending on the referent, can be indexed by a pronominal clitic on the verb (*=nyo* ‘3SG’ in AV and *ku=* ‘1SG’ or *mu=* ‘2SG’ in PV). The Secondary Argument frequently forms a tight association with the verb, disallowing any intervening elements and forming a constituent that we refer to as the predicate complex.

²Given that bare verbal transitive roots almost exclusively mark for PV, they will be glossed in the examples as PV.

- (4) *antakan ni nyak ng-gulai khetak telung,*
 before DEM.ADDR 1SG AV-soup long.bean eggplant
 ‘yesterday, I made long bean soup with eggplant,’
 (BJM02-048-02, 00:05:20–00:05:22, Speaker: Ita)
- (5) *sawah wo Ros ni kak di-akuk=nyo lih Satimo,*
 field sister R. DEM.ADDR already PV-take=3SG by S.
 ‘Ros’s field has already been bought by Satimo,’
 (BJM02-005, 00:35:36–00:35:38, Speaker: Ari)

The suffixes *-kun* ‘CAUS/BEN’ and *-i* ‘LOC/PLUR’ most often function as valency increasing affixes that have both causative and applicative functions (see McDonnell & Truong 2024a). The use of one of these suffixes allows, for example, for typically intransitive verbal roots to become transitive or for monotransitive roots to become ditransitive with benefactive, instrumental, or goal applied arguments. They are also used to derive transitive verbs from noun roots and even trigger semantic changes on the verb, such as iterative and pluractional meanings (see Truong & McDonnell 2022).

Aside from the fact that the Secondary Argument often forms a predicate complex with the verb, word order in Nasal is otherwise variable, as is the case with many other languages of Western Indonesia. Argument fronting frequently occurs for emphasis regardless of voice construction. Because of this, both the marked word order, in which the Primary Argument follows the predicate complex (6), and the unmarked order, in which the Primary Argument precedes the predicate complex (7), are common.

- (6) *pakai sintikh udi nyak*
 use flashlight DEM.DIST 1SG
 ‘I used that flashlight.’
 (BJM02-012, 00:37:18–00:37:19, Speaker: Een)
- (7) *nyak ng-ucap-kun tekhimo kasih lawan kai,*
 1SG AV-say-CAUS/BEN thank you with 2PL
 ‘I say thank you to you all.’
 (BJM02-056-01, 00:13:11–00:13:13, Speaker: Arifin)

2.2 Corpus of conversational data

This study reports on the use of fillers in six transcribed conversations – totalling approximately eight hours of speech – taken to be representative of both men

and women ranging from 20 to 60 years old from a larger corpus of 24 conversations. We revisit this larger corpus in §6. These conversations were coded for the various uses of placeholders and hesitators in Nasal along with other relevant information regarding repair, disfluencies, and syntactic position, among others. This coding allows us to take a comprehensive look at the uses of fillers in Nasal (without cherry picking), reporting on the relative frequencies of fillers and their functions. These recordings included both audio and video, but while the video assisted some in determining a given word's function (for example, pointing to indicate a prototypical demonstrative function), syntactic and prosodic cues were most reliable in determining the role of a placeholder and, thus, coding only applied to recorded speech. A brief description of each of the six conversations is provided in Table 1, and the complete schema used for coding this dataset is shown in Appendix A. The remainder of the coded examples cited in this paper are taken from these six conversations.

Table 1: Conversations included in the corpus.

ID	Description
BJM02-012	A conversation between two cousins – Een (F, 23) and Nera (F, 29) – and their friend Metia (F, 26) in Gedung Menung about their families, weddings and marriages, and the weather
BJM02-014	A conversation between three friends – Bahrul (M, 39), Herman-syah (M, 37), and Karnain (M, 45) – in Tanjung Baru primarily about fishing and keeping livestock
BJM02-022	A conversation between three friends – Yogi (M, 26), Johan (M, 57), and Sumardianto (M, 35) – in Tanjung Betuah about finding work, growing various plants, and the prices of things
BJM02-044	A conversation between three friends – Susi (F, 36), Seni (F, 37), and Mihar (F, 42) – in front of a house in Gedung Menung about planning going to the beach, cooking, looking for snails, and selling things online
BJM02-050	A conversation between three friends – Surnila (F, 50), Ros (F, 35), and Resta (F, 31) – in Tanjung Betuah about family happenings, cooking and food, and using technology
BJM02-058	A conversation between three friends – Bahuri (M, 51), Redo (M, 29), and Wawan (M, 36) – in Tanjung Betuah discussing Nasal history and stories about Nasal ancestors

3 Overview of fillers in Nasal

Unlike some other languages of Western Indonesia, Nasal does not have a dedicated filler (compare, for example, Karo Batak *kadih* in Woollams 1996: 118; South Barisan Malay *anu* in McDonnell & Billings 2025 [this volume]). Instead, demonstrative and interrogative pronouns frequently fulfill this role and are variously employed as both hesitators and placeholders.

Nasal employs a three-way contrast between demonstratives for indicating objects near the speaker (*ajo* ‘DEM.SP’), near the addressee (*ani/heni* ‘DEM.ADDR’), and far from both the speaker and addressee (*udi* ‘DEM.DIST’). Each of these can variously be used as a demonstrative pronoun and as a demonstrative determiner, which is the last element in an NP. Such a three-way contrast is common in Austronesian languages, but, as noted above, the complexity of the entire demonstrative system as a whole is rather remarkable (see Table 2 below). All three demonstratives have an emphatic form with initial *ng-* as well as a determiner form that occurs at the right edge of a noun phrase and is otherwise semantically equivalent to the determiner uses of the independent forms. This demonstrative series also forms the roots for locative and manner demonstratives. Although Nasal has a dedicated set of manner adverbs based on these roots, the emphatic forms with initial *ng-* are also frequently utilized in this role. An overview of this demonstrative system is provided in Table 2, and examples of pronominal and determiner use of demonstratives are provided in (8) and (9) below, respectively. In these and following examples, an arrow (→) is used to indicate the line in which the referenced demonstrative, interrogative, or filler is used.

Table 2: Nasal demonstrative system.

	Ind.	Emph.	Det.	Loc.	Manner
DEM.SP	<i>ajo</i>	<i>ngajo</i>	<i>sijo</i>	(<i>d</i>) <i>ijo</i>	<i>jeujo</i>
DEM.ADDR	<i>ani/heni</i>	<i>ngani</i>	<i>ni</i>	(<i>d</i>) <i>isan</i>	<i>jeusan</i>
DEM.DIST	<i>udi</i>	<i>ngudi</i>	<i>sudi</i>	(<i>d</i>) <i>udi</i>	<i>jeudi</i>

- (8) 1 S: *na;*
 well
 ‘ah,’
 2 → ***ajo*** *khetuk*
 DEM.SP delicious
 ‘these are delicious.’

(BJM02-044-01, 00:09:56–00:09:57, Speaker: Seni)

- (9) 1 S: *mak nihan njuk=nyo:*
NEG very PV.give=3SG
'she really didn't give (it to her).'
2 *mak njuk=nyo:*,
NEG PV.give=3SG
'she didn't give (to her);'
3 → *suluh sudi*
firewood DEM.DIST
'that firewood.'

(BJM02-044-01, 00:13:39–00:13:43, Speaker: Susi)

All three of the demonstrative pronouns can operate as both hesitators and placeholders. Although the near-speaker demonstrative *ajo* is not found to be used as a hesitator in the corpus, its absence may simply be due to its lower frequency as a filler overall (as reflected by only a single clear use as a placeholder being found in the corpus) rather than an inability of the near-speaker demonstrative to be used as a hesitator. This stands in contrast to Indonesian (Wouk 2005) and Besemah (McDonnell & Billings 2025 [this volume]) in which the proximal demonstrative is much more frequent.

In addition to the demonstrative pronouns, the interrogative pronoun *api* 'what' is also frequently found operating as a filler and, as with the demonstratives, is used as both a hesitator and a placeholder. In its typical non-filler functions, *api* 'what' is used as an interrogative pronoun (10), an indefinite pronoun (11, 12), and an alternative correlative conjunction (13). The other interrogatives – *sapo* 'who', *kebilo* 'when', (*d*)*ipo* 'where', *jeupo* 'how', *sipo* 'which' – are not used as fillers. In theory, the interrogative *ngapi* 'why' could be homophonous with the morphologically complex *ng-api* 'Av-what', although it is not attested in our corpus.

- (10) *n-(t)akhuk api tian?*
AV-plant what 3PL
'what did they plant?'

(BJM02-012-01, 00:43:23–00:43:24, Speaker: Metia)

- (11) *sekh:ah ngio ng-(k)icik a:pi*
surrender EMPH AV-say what
'it's up to us, (we can) talk about whatever'

(BJM02-012-01, 00:18:14–00:18:16, Speaker: Een)

- (12) *wat api-api untuk ayuk*
 exist what for older.sibling
 ‘there’s something for him.’

(BJM02-046-01, 00:18:03–00:18:05, Speaker: Lili)

- (13) *sebenakhnyo husnul api khusnul?*
 actually h. what k.
 ‘actually (how is it pronounced), *husnul* or *khusnul*?³

(BJM02-012-01, 01:05:27–01:05:30, Speaker: Een)

Although Nasal demonstratives and the interrogative *api* operating in their prototypical functions are unable to take affixes or proclitics, both of these types of fillers can appear with a variety of verbal and nominal affixes as well as the pronominal proclitics *ku*=‘1SG’ and *mu*=‘2SG’. Since hesitators by definition do not occur with any morphological marking, morphologically complex forms are entirely restricted to the placeholder use of the fillers. As opposed to affixes and proclitics, enclitics are found both when demonstratives and *api* are used in their typical function and when used as a placeholder.

For the remainder of this chapter, general claims made about hesitators or placeholders, unless otherwise specified, should be taken in reference to all four function words as fillers under investigation, both demonstrative and interrogative. Other general distributional tendencies or formal differences between any of the function words when used as fillers not mentioned here were not apparent in our coding and require additional data for analysis.

3.1 The indeterminacy of placeholders and hesitators

Given their polyfunctional nature, it is often difficult to determine whether a given filler in Nasal is operating as a hesitator or as a placeholder. This is made even more difficult by the fact that the typical diagnostics for distinguishing between these two filler types – such as case, agreement, and other morphological marking – are largely absent in Nasal. The only help in this matter comes from voice, applicative, and nominalizing affixes, although even these are not always obligatory; see, for example, (14) and (15), both of which lack morphology in the placeholder but fill the syntactic position of a noun and a verb respectively. Even clitics are not of any immediate help. While first- and second-person enclitics only attach to these function words when they are used as placeholders, they

³In this example, the speaker is asking how to pronounce an Arabic word in the the phrase *husnul khatimah* ‘a beautiful conclusion’.

only rarely attach to fillers (see Table 8 further below). Although similarly disambiguating placeholders from hesitators, the third-person enclitic (as used in 16), on the other hand, does not necessarily disambiguate filler from non-filler uses of function words since it can freely attach to function words to mark definiteness. In (16), the repair provides clear evidence of its placeholder status.

- (14) 1 →B: *duwik .. ani:*
many DEM.ADDR
'(there are) many whatchamacallit.'
2 *anak .. api kio?*
child what EMPH
'baby .. what is it?'
3 *ipun hukhang*
baby shrimp
'baby shrimp.'
(BJM02-014-01, 00:49:39–00:49:43, Speaker: Baidi)
- (15) 1 →J: *nyo sudi hago ani:*
3SG DEM.DIST want DEM.ADDR
'he wants whatchamacallit.'
2 *be-variasi:*
MID-variety
'(for it) to have variety.'
3 *pajuhan=nyo sudi*
food=3SG DEM.DIST.
'his food.'
(BJM02-014-01, 01:01:56–01:02:00, Speaker: Edwar)
- (16) 1 A: *pada saat ani,*
at time DEM.ADDR
'at that time,'
2 → *mimang wat api=nyo,*
really exist what=3SG
'there really was whatchamallit,'
3 *sejarah=nyo*
history=3SG
'a story.'
(BJM02-058-01, 00:11:04–00:11:09, Speaker: Ahmad)

As with many roots in Nasal, however, it is not always possible to distinguish the word class of the filler, whether demonstrative or interrogative pronoun. That is, without specific morphological marking on the affixed word indicating one class or the other (for example, the verbal AV prefix *N-* or the nominalizing suffix *-an*), a root can operate in its normal function as a demonstrative/interrogative, as a placeholder noun, or as a placeholder intransitive verb (for placeholder transitive verbs, the valency-increasing applicative *-kun* is present in virtually every case, as will be seen later).

Two additional factors complicate matters. As mentioned in §2, Nasal, like other languages of Western Indonesia, tends to omit arguments without any indexing on the verb. When arguments have been omitted, an otherwise evident verbal placeholder can appear very similar to a hesitator or, at times, a prototypical demonstrative use, given that all three can utilize a bare root. Second, placeholders in Nasal frequently are not repaired. In the absence of a word search or recycling, the placeholder can again look very similar to a demonstrative pronoun. Of the coded data, more than a third of placeholder fillers are not repaired (see §5.1).

For the examples given above (and many others in the corpus), the lack of preceding context for the intended meaning of *ani* ‘DEM.ADDR’ – whether verbal or gestural – in (14) and (15) eliminates the possibility of it being used anaphorically. The disambiguation of hesitator and placeholder role for these two examples can be found in the repair, although this does not always occur. In cases lacking repair, participation in the syntax of the clause and continuation of normal intonation are key markers for placeholder uses. For example, in (16) above, *api=nyo* could be, as is often the case, a self-addressed question. However, the participation in the clause as Primary Argument, the continuing intonation, and the recycled *=nyo* all point to the interrogative being used as a placeholder. Such contextual, syntactic, and intonational clues allow us to reliably code many of the hesitator and placeholder uses of fillers, but a large number remain indeterminate.

Table 3 contains frequency counts from the six-conversation corpus (approximately 60,000 words) for both the filler and prototypical uses of each function word. While the 18 speakers varied in their use of hesitators and fillers, there were no clear discrepancies or inconsistencies in the use of each across speakers in our sample. The demonstrative *udi* ‘DEM.DIST’ as a hesitator (6 tokens) and the interrogative pronoun as a placeholder (10 tokens) were each only used by a single speaker. However, these uses were checked against additional participants and conversations in the larger corpus and it was found that many additional speakers outside of these six conversations use *udi* ‘DEM.DIST’ and the interrogative

in the same functions. Thus, these numbers are likely a result of low frequency, and not simply speaker preference.

Demonstrative and interrogative pronouns more frequently function in their prototypical demonstrative or interrogative use. As fillers, the demonstratives have a clear preference to be placeholders, and the interrogative has a preference to be a hesitator. Each of these uses is discussed further in the following sections. Rather than forcing a particular analysis, a filler was coded as indeterminate where either its role as a placeholder or hesitator was unclear or whether it should be considered a filler at all was unclear. The number of these cases is also shown in Table 3.

Table 3: Hesitator and placeholder frequencies of Nasal fillers.

	Filler			
	Hesitator	Placeholder	Indeterminate	Prototypical Use
DEM.SP	-	1	1	237
DEM.ADDR	24	77	27	496
DEM.DIST	6	18	3	289
Interrogative	22	10	2	546

4 Hesitator fillers

As seen in Table 3, the filler most likely to be used as a hesitator rather than a placeholder is the interrogative pronoun *api* ‘what’, as in (17).

- (17) 1 N: *Bakakh m-ulang jenu*
 B. INTR-return earlier
 ‘Bakakh returned then.’
- 2 → *angan=ku hago ng-(k)ikit di .. api disan,*
 thought=1SG want AV-fish at what there
 ‘I thought he wanted to fish at, uh, there,’
- 3 *di Saudi tian ni*
 at S. 3PL DEM.ADDR
 ‘where Saudi and them were.’
- (BJM02-014-01, 00:24:16–00:24:20, Speaker: Nain)

In this example, Nain uses *api* in line 2 to hold the floor when retelling how Bakakh had to return home to get his fishing net. However, when used to express hesitation, *api* is much more often encountered in self-addressed questions (63 instances in the 6-conversation corpus, not included in the frequency counts of Table 3) compared to *api* on its own (22, as shown in Table 3). In (18), for example, Ahmad is about to recount a story and uses the phrase *api sudi* ‘what’s that?’ in line 2 to delay while remembering the story’s content and main character.

- (18) 1 A: *amun: cekhito;*
if story
'if it's a story,'

2 → *yang api sudi,*
REL what DEM.DIST
'that, what's it,'

3 *yang pintakh,*
REL smart
'that (is about somebody who's) smart,'

4 *amun cekhito Arab sudi kan?*
if story Arab DEM.DIST TAG
'if it's one of those Arab stories, right?'

5 *ido Abu Nawas:*
true A. N.
'then that (is going to be one about) Abu Nawas.'

Similarly, in (19), Nera is telling what she had recently cooked and uses the phrase *api kio namonyo* ‘what’s it called?’ while trying to recall in line 2. Many such similar self-addressed questions in which the common element is always *api* are frequently used to express hesitation.

- (19) 1 N: *kito*—
 1PL.INCL
 ‘we’
 2 → *api kio namo=nyo?*
 what EMPH name=3SG
 ‘what’s it called?’

- 3 *ny-(s)ani:k,*
 AV-make
 ‘made,’
4 (*TSK*) *gulai talus sudi*
 soup taro DEM.DIST
 ‘that taro soup.’
 (BJM02-012-01, 00:09:54–00:09:59, Speaker: Nera)

The distinction between a hesitator use of *api* and a self-addressed question appears on the surface to be clear-cut. The former is a non-referential interjection vocalizing hesitation in speech while the speaker attempts to formulate or complete an utterance. The latter is a fully-formed and referential interrogative, externally expressing the speaker’s uncertainty during an internal recollection of a difficult to remember or lost referent. However, as with other matters involving fillers, distinguishing these two in practice is difficult. First, both the hesitator and the self-addressed question may be a single word or an entire phrase. For interjective hesitators, both *api* and *api kio* (with the emphatic discourse marker *kio* ‘EMPH’) are often employed. For self-addressed questions, the same can be said, with *api?* ‘what?’, *api kio?* ‘what?’, *api kio gelakhnyo?* ‘what’s its name?’, or *api namonyo?* ‘what’s its name’, all being common choices. Thus, unlike English ‘what do you call it?’ and ‘whatchamacallit’ which are distinguished phonologically (Enfield 2003), the structure of the hesitator utterance in Nasal does not immediately help in distinguishing the two.

Second, while there are some instances of clear-cut intonational differences between an interjective hesitator *api* and a self-addressed question, this is not always the case. Other non-lexical markers, if they occur, may help distinguish the two, such as lengthening or truncation of the preceding IU leading into a self-addressed question. In Figure 1 and Figure 2 we see two such distinguishable cases.

In Figure 1, Edwar is talking about people from a certain area but forgets the name of the area, first using *ani* as a placeholder and then hesitating while trying to remember the name. This attempt at recall is expressed with the self-addressed question *api gelakh ni?* ‘what’s this called?’ where *ani* appears yet again in a reduced form *ni* to reference the sought referent. That this final element *ni* of the IU immediately preceding *api* is lengthened, an independent IU is started, and distinctly interrogative intonation is used all indicate that this is a self-addressed question. In Figure 2, Ahmad is telling a story in which somebody catches a deer. He starts saying the wrong word for deer (*kijang*), cuts himself off, hesitates

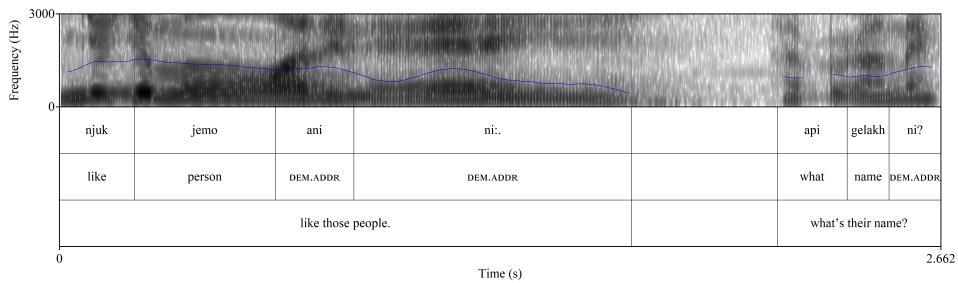


Figure 1: Self-addressed question *api gelakh ni?* used for hesitation (BJM02-014-01, 01:07:24–01:07:27, Speaker: Edwar).

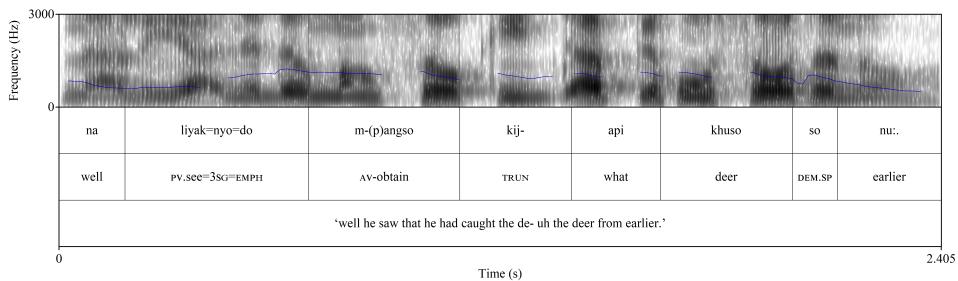


Figure 2: Interjective hesitator *api* in continuous speech (BJM02-058-01, 00:14:38–00:14:41, Speaker: Ahmad).

briefly with an interjective *api*, and then continues normally. No independent IU is started, intonation is continuous over the whole utterance, and truncation occurs, but only as a means to prevent production of the wrong word.

While such distinguishable cases do exist, many are ambiguous for an interjective hesitator or a self-addressed question, contributing further to the indeterminacy in establishing the filler uses of function words in Nasal.

The demonstratives *ani/heni* ‘DEM.ADDR’ (20) and *udi* ‘DEM.DIST’ (21) are also often used to signal hesitation. In (20), Edwar is talking about how it would be easier to use fishing nets with two people and uses *ani* ‘DEM.ADDR’ as a hesistative in line 3 when he cannot remember the location of where the second person would be standing. This hesitation is immediately followed up by a self-addressed question to hold the floor within the same IU.

- (20) 1 E: *nyo sudi kan*,
3SG DEM.DIST TAG
'he, right,'

- 2 *khampusan=nyo ni,*
 shallows=3SG DEM.ADDR
 '(can go into) these shallows,'
- 3 → *khamp- ani:*,
 DEM.ADDR
 'uh,'
- 4 *api gelakh ni?*
 what name DEM.ADDR
 'what's it called?'
- 5 *khelum pinggikh*
 deep edge
 '(go to the part that's) deep at the edge.'
- (BJM02-014-01, 00:21:35–00:21:41, Speaker: Edwar)

Johan's use of *udi* multiple times in (21) in lines 3 and 4 shows his hesitation in trying to find the right description for a kind of bread eaten during war.

- (21) 1 J: *khuti so duwik telul=nyo susu=nyo*
 bread DEM.SP many egg=3SG milk=3SG
 'this bread has a lot of eggs and milk.'
- 2 *e- kung ngam-*
 not.yet
 'not yet—'
- 3 → *udi khuti hulun,*
 DEM.DIST bread person
 'uh, the bread that people,'
- 4 → *pe- udi,*
 DEM.DIST
 'uh'
- 5 *pekhang sudi kidah,*
 war DEM.DIST EMPH
 'those ones (that you eat) in war'
- 6 *mbak i:jo nihan*
 same.size DEM.SP very
 'it's the same size as this.'
- (BJM02-022-01, 01:09:48–01:09:55, Speaker: Johan)

While it is most common to encounter hesitators on their own, other markers of hesitation (e.g. pause, truncation, lengthening) are frequently used in conjunction with them. In fact, in languages like Nasal where hesitator particles akin to English *uh* and *um* are far less frequent (although not entirely absent, as with *e* in (23) further below), it is more common to find these additional means of expressing hesitation, as has also been noted, for example, for Ilokano and Lauje (Streeck 1996) as well as Tagalog (Himmelmann 2014). Those that were identified in the coding are shown in Table 4. In some cases, more than one of these co-markers of hesitation were found to co-occur with the hesitator.

Table 4: Co-markers of hesitation.

	None	Pause	Truncation	Lengthening	Glottalization
Demonstrative	16	9	7	2	1
Interrogative	14	2	6	-	-

The most frequent co-marker of hesitation is a brief pause (17) or truncation of a lexical item or IU (20, 21). Lengthening of the word immediately preceding the hesitator (19) is also present but not common, and we identified one case where hesitation is signalled additionally by glottalization (22). In this last example (22), Johan and Yogi are discussing finding work in Nasal. When Johan asks what work Yogi has found in Nasal, Johan expresses hesitation in finishing his response through both glottalization and the demonstrative *ani*.

- (22) 1 J: *api lukak di dijo so?*
 what work at here DEM.SP?
 ‘what work is there (for you) here?’
- 2 Y: *mak deduk igo;*
 NEG exist too.much
 ‘there’s not too much,’
- 3 → *amun wat .. %lukak .. %ani %tukhu:k*
 if exist work DEM.ADDR join.in
 ‘if there is .. work .. uh, I do it.’
- (BJM02-022-01, 00:04:02–00:04:07, Speakers: Johan, Yogi)

Proportionate to their frequency of use, there do not appear to be any major differences between the uses of these non-lexical co-markers with demonstratives

or interrogatives. Although a brief pause is seemingly more commonly found before a demonstrative used as a hesitator, a larger sample would need to be taken in order to confirm this correlation.

Most other methods of signalling hesitation in Nasal that can be used independently of lexical hesitators are either interjective (as with *na* ‘well’ in (24) or *e* ‘um’ in (26) below) or self-addressed questions (see above). Questions containing *api* – as in (18) and (19) above – remain the most common of these, even more so than the use of a demonstrative or *api* on its own.

5 Placeholder fillers

Placeholders are often described in reference to their target (e.g. word class, grammatical relation), which according to Podlesskaya (2010) refers to the word (or presumably phrase) that the speaker intended. One issue with the term *target*, however, is that it requires the analyst to know what the speaker intended, which is only truly possible when the placeholder has been repaired. Since the majority of forms are not repaired in Nasal, we follow Fox (2010), who frames her discussion of fillers using an Interactional Linguistics approach. She describes placeholders as “fulfill[ing] the syntactic projection of the turn so far” (Fox 2010: 2). By describing fillers in this way, we need not make any claim about the word or phrase the speaker intended in the absence of repair. Instead, we simply describe the syntactic position of the placeholder. In practice, however, fulfilling a syntactic projection and target can be considered rough equivalents with the caveat that we make no claim about what the speaker intended.

In the six conversations analyzed, there are a total of 106 placeholder uses of fillers. As mentioned in §4, there is a tendency for demonstratives in Nasal to be used as placeholders rather than as hesitators (see Table 3). The most common demonstrative to be used as a placeholder (as with hesitators) is *ani/heni* ‘DEM.ADDR’ (23), followed then by *udi* ‘DEM.DIST’ (24), and finally *ajo* ‘DEM.SP’ (25). Metia uses *ani* in line 1 of (23) as a placeholder when recommending that Een buy an album for her wedding photos. There, *ani* fulfills the syntactic projection of a clausal argument, and the target for the placeholder is *bingkai foto* ‘picture frame’, as demonstrated by the repair in line 2.

- (23) 1 →M: *beli-kun pai e ani wu:i*
buy-CAUS/BEN first um DEM.ADDR EMPH
'buy, uh, a whatchamacallit please.'

- 2 *bingkai foto*
 frame photo
 ‘a picture frame’
 (BJM02-012-01, 00:58:43–00:58:48, Speaker: Metia)

In (24), Johan is talking about what he had recently planted and uses *udi* to hold the place while he recalls *kacang buncis* ‘green beans’. In this case, the placeholder *udi* fulfills the syntactic projection of a nominal argument.

- (24) 1 →J: *na n-(t)akhuk .. udi:* *nyak,*
 well AV-plant DEM.DIST 1SG
 ‘well, I plant whatchamacallit,’
 2 *api?*
 what
 ‘what?’
 3 *kacang buncis*
 green bean
 ‘green beans.’
 (BJM02-022-01, 00:57:30–00:57:34, Speaker: Johan)

Finally, Edwar uses both *ajo* and *ani* in (25) for somebody he had seen earlier, while he tries to remember his name, fulfilling a nominal projection in both instances.

- (25) 1 E: *pikho=do m-beli*
 how.much=EMPH AV-buy
 ‘however much (it is, he will) buy (it),’
 2 → *d- .. d- dang ajo*
 mister DEM.SP
 ‘Mr. whatchamacallit.’
 3 → *dang: ani .. Brad jenu*
 mister DEM.ADDR B. earlier
 ‘Mr. whatchamacallit .. Brad from earlier.’
 (BJM02-014-01, 01:10:28–01:10:32, Speaker: Edwar)

Unlike hesitators, placeholders may take morphological marking both via affixation and cliticization. Affixation has a clear tendency to occur when the placeholder is operating in a verbal function, as in (26), accounting for 18 of the 20 affixed forms. The remaining two are nominal placeholders, as in (27). In (26), Susi

and Seni are talking about somebody else's kid being sick. Susi asks about how to treat the sickness, and Seni describes the process, using *di-udi-kun* to hold the place of an unrepaired verb, understood from context to be rubbing on oil.

- (26) 1 SU: *a:pi ubat=nyo?*
DEM.INT medicine=3SG
'what's the treatment for it?'
- 2 SE: *e,*
HES
'uh'
- 3 *yang panas-pana:s*
REL RDP-hot
'that hot stuff.'
- 4 → *di-udi-kun minyak kayu puti:h,*
PV-DEM.ADDR-CAUS/BEN oil cajuput
'cajuput oil is whatchamacallit-ed (rubbed on)'
- 5 *adu tapal-i busung, ...*
after PV.apply-LOC/DIST belly
'and then it's applied to the belly'
(BJM02-044-02, 00:04:34–00:04:40, Speakers: Susi, Seni)

Edwar uses the nominalizing suffix *-an* in *ani-an* 'DEM.ADDR-NMLZ' in the place of 'difference' in (27) while remarking that there's no difference between the two types of bamboo he and Nain are discussing. This interpretation is reinforced by his following IU.

- (27) 1 N: *bet:i:k m-belah=nyo,*
nice AV-cut=3SG
'it's easy to split it,'
- 2 E: *beti:k*
nice
'easy.'
- 3 → *mak deduk [ani-an]*
NEG exist DEM.ADDR-NMLZ
'there isn't any whatchamacallit (=difference).'
- 4 N: *[m-belah][₂=nyo]*
AV-cut=3SG
'(it's easy to) split it.'

- 5 E: [*z samo*] *gawuh*
 same just
 ‘they’re just the same.’
 (BJM02-014-01, 00:37:25–00:37:29, Speakers: Nain, Edwar)

In the data considered here, cliticization on a placeholder is only encountered with the enclitic =*nyo* ‘3SG’ and primarily among nominals (seven occurrences as in (28)) rather than verbals (two occurrences as in (29)). This reflects the more common usage of =*nyo* ‘3SG’ as a possessive marker or marker of definiteness rather than as an indexed argument in transitive verb constructions. Baidi uses *heni=nyo* ‘DEM.ADDR=3s’ in (28) when discussing the necessity of clipping the wings of domesticated ducks. The placeholder *heni* ‘DEM.ADDR’ holds the place for *kepi* ‘wing’ while Baidi thinks of the right word, =*nyo* here referring to the ducks.

- (28) 1 B: *amun kak se=khatus*,
 if already one=hundred
 ‘if (you’ve) already (done) one hundred (of them),’
 2 *kak n-(t)anggung*,
 already AV-tiring
 ‘it’s already tiring,’
 3 → *n-(t)etuk-i heni=nyo*,
 AV-cut-LOC/DIST DEM.ADDR=3SG
 ‘cutting off **their whatchamacallits**,’
 4 *kepi=nyo*
 wing=3SG
 ‘their wings.’
 (BJM02-014-01, 01:02:09–01:02:11, Speaker: Baidi)

In (29), Gadis is joking about Resta looking sleepy on the recording. She uses *ng-udi-kun* to hold the place for a verb referring to watching and =*nyo* to refer to the recording itself.

- (29) 1 G: *betik amun dengi*
 nice if night
 ‘it’s nice at night.’
 2 *ai*,
 ah
 ‘ah,’

- 3 *khaso=ku kepayaha:n yang se=bigi sudi la=nyo,*
 feel=1SG tired REL one=CLASS DEM.DIST say=3SG
 'I feel like he's going to say, that one's tired,'
- 4 → *keliyakan .. nti tian ng-udi-kun=nyo*
 seem later 3PL AV-DEM.DIST-CAUS/BEN=3SG
 'looking .. later when they whatchamacallit (=watch the video recording).'
- (BJM02-050-03, 00:03:03–00:03:08, Speaker: Gadis)

The following subsections describe how these placeholders operate both in repair sequences and the syntactic positions they occupy (i.e. the syntactic projections they fulfill).

5.1 Repair

As seen in Table 5, more than a third of all cases of placeholder fillers are not repaired. There appears to be no clear correlation between repair and the form of the placeholder that would not be otherwise attributable to general distribution of the fillers. We first consider repaired placeholders.

Table 5: Placeholder frequencies of Nasal fillers by repair.

	Repaired	Unrepaired
DEM.SP	1	-
DEM.ADDR	47	30
DEM.DIST	14	4
Interrogative	8	2
Total: 106 =	70	36

5.1.1 Repaired placeholders

As has been noted cross-linguistically (Németh 2012), the primary repair strategy in Nasal is self-initiated self-repair, but both other-initiated and other-repair are attested in our corpus, as in (30) and (31) below. When Ahmad cannot remember the word he used for 'gang' in his retelling of a Nasal story in (30), Johan helps repair the filler *api* in line 4.

- (30) 1 A: *habis-*
 gone
 ‘gone—’
 2 *a,*
 ah
 ‘uh,’
 3 *belo=do yang,*
 gone=EMPH REL
 ‘they were gone,’
 4 → *api so nu,*
 what this earlier
 ‘these whatchamacallit I just mentioned,’
 5 J: *gekhumbulan so nu*
 gang this earlier
 ‘the gang.’
 6 A: *gekhumbulan*
 gang
 ‘the gang.’
 (BJM02-058-01, 00:07:21–00:07:25, Speakers: Ahmad, Johan)

Same-turn self-repair in Nasal typically occurs in the immediately following IU. Neither the choice of filler nor the syntactic position of the placeholder appears to have any correlation with what material, if any, is recycled during repair. Word searches are not altogether infrequent but rarely involve participation from other participants. One of the few cases of this can be seen in (31) where Baidi assists Edwar in trying to recall the name for a specific kind of fish.

- (31) 1 →E: *nju:k ani [api?]*
 like DEM.ADDR what
 ‘(it’s) like whatchamacallit, what?’
 2 B: *[iwo bela]to?*
 fish b.
 ‘*belato* fish?’
 3 E: *ayin=nyo iwo belato*
 NEG=3SG fish kind of fish
 ‘it’s not *belato* fish.’

- 4 *anak-anak tungkul*
 child-RDP mackerel
 ‘young mackerel.’
- 5 *njuk anak tungkul ni*
 like child t. DEM.SP
 ‘it’s like young mackerel.’
- 6 *tapi [%gemuk]*
 but fat
 ‘but fat.’
- 7 B: [*o iwo dinci*][_{2s}]
 oh fish sardine
 ‘oh, sardine’
- 8 E: [₂*o*]_o,
 yeah
 ‘yeah,’
- 9 *mungkin=do ani=do*
 maybe=EMPH DEM.ADDR=EMPH
 ‘maybe that’s it.’
- (BJM02-014-01, 00:27:49–00:27:59, Speakers: Edwar, Baidi)

The majority of word searches, however, involve only the speaker who utilized the placeholder, as in (32) and (33). Yogi searches for the right word for applying poison to protect his plants in (32), and Johan searches for the right word for a type of banana whose seedlings somebody had recently picked for planting in (33).

- (32) 1 J: (%) *adak beli-kun* *tiodan*
 or pv.buy-CAUS/BEN poison
 ‘or buy poison.’
- 2 Y: *oo*
 yeah
 ‘yeah.’
- 3 J: [*sempekhut tiodan*]
 spray poison
 ‘(you can) spray the poison.’

- 4 → Y: *[ani-kun] nihjan*
 PV.DEM.ADDR-CAUS/BEN very
 '(you can) whatchamacallit.'
- 5 *pik-kun,*
 PV.put-CAUS/BEN
 'put it,'
- 6 *suntik-kun*
 PV.inject-CAUS/BEN
 'inject it'

(BJM02-022-01, 00:25:48–00:25:54, Speakers: Johan, Yogi)

- (33) 1 J: *dijo so kak .. akuk-i Tris anak=nyo*
 here DEM.SP already PV.take-LOC/DIST T. child=3SG
 'here Tris's mom already took their seedlings (lit. their children).'

- 2 *Ambon*
 kind of banana
 'Ambon (bananas).'
- 3 → *khan udi,*
 and DEM.DIST
 'and whatchamacallit,'

- 4 *api gelakh=nyo:?*
 what name=3SG
 'what's it called?'

- 5 *ajo:,*
 DEM.SP
 'whatchamacallit,'

- 6 *pisang huwai:,*
 kind of banana
 'huwai bananas,'

- 7 *jantan*
 kind of banana
 'jantan bananas.'

(BJM02-022-01, 00:23:17–00:23:24, Speaker: Johan)

If a clitic is used in conjunction with a placeholder, it is more often found recycled also in the repair, whether in reference to a verbal or a nominal, as in (28)

above, but, again, this is not obligatory. Similarly, if a verb is repaired, typically both the voice and transitivity affixes of the placeholder will be recycled, as in (32). However, given the hesitative nature of placeholders, this, too, is not always the case, as discussed further below with (38) and (39).

5.1.2 Unrepaired placeholders

In more than a third of the coded placeholders, no explicit repair is made and conversation continues without any apparent issue, as in (34) where Edwar's placeholder use of the second *ani* in line 2 is understood by Baidi and Nain from its contrast with the referential *ani* (referring to a previously mentioned 'dragnet') immediately preceding it in line 1.

- (34) 1 E: *man mak hago ani*,
if NEG want DEM.ADDR
'if you don't want that one (=dragnet),'
2 → *m-beli=do ani*,
AV-buy=EMPH DEM.ADDR
'buy a whatchamacallit (=different kind of net),'
3 *api gelakh=nyo ni?*
what name=3SG DEM.SP
'what's this one called?'
4 B: *genggung gawuh Manulah madah*
bring just M. let's
'let's just bring it to Manulah.'
5 N: *incang-kun gawuh jeijo*
PV.raise-CAUS/BEN just like.that
'we can fish just like that.'
(BJM02-014-01, 00:17:04–00:17:11, Speakers: Edwar, Baidi, Nain)

While in some cases placeholders are used for taboo reasons and thus not repaired, as in both (35) and (36) where *ani* is used to refer to dying, the majority of cases where they are not repaired appear to simply be understood from context, as in (26) above.⁴ The use of a placeholder for taboo reasons does not appear to have any marked differences with other placeholder uses that are not repaired.

⁴In all three examples of avoidance in our coding, *ani* is the filler used. However, this simply could be a result of frequency, and we suspect that other fillers are able to fulfill the same function.

- (35) 1 →E: *semenjak ani pekdang so,*
 since DEM.ADDR uncle DEM.SP
 ‘since my uncle whatchamacallit (=passed away),’
 2 *mak pekhenah .. ngekhe- ny-(c)akak-kun henil-henilan*
 NEG ever AV-raise-CAUS/BEN rancid
agi
 again
 ‘I haven’t brought any more rancid things (=fish) again.’
 (BJM02-014-01, 00:28:30–00:28:34, Speaker: Edwar)
- (36) 1 →R: *ilmu=nyo sudi kak bel:o masih jugo kung ani*
 magic=3SG that already used.up still also not.yet DEM.ADDR
 ‘her magic is already gone, but she also still hasn’t yet
whatchamacallit (=passed away).’
 2 *kak pacul-kun*
 already PV.free-CAUS/BEN
 ‘she’s already been freed.’
 (BJM02-050-01, 00:20:17–00:20:21, Speaker: Resta)

5.2 Syntactic position

Of the 106 instances of placeholders, more than two thirds occur in the syntactic position (i.e. fulfill the syntactic projection) of nominals, either as arguments, objects of prepositions, or nominal modifiers (e.g. possessors). This is consistent with the cross-linguistic tendency for placeholders to occur in nominal positions (Podlesskaya 2010). Although, as discussed previously, demonstrative and interrogative forms do not necessarily require marking for particular word classes, it is likely their origin in referential and pronominal functions that reflects this greater trend to take the place of nominals. Nominal placeholders rarely take any morphology (as in, for example, (27) above), and such morphology is by no means obligatory.

When a placeholder fulfills the syntactic projection of an intransitive verb, explicit morphology is equally likely as being morphologically unmarked (4/8). It does, however, occur in virtually every instance of a placeholder fulfilling the projection of a transitive verb (11/12), as in (37).⁵ In this sequence, Edwar asks

⁵Placeholders fulfill intransitive and transitive projections 20 times in the corpus while placeholders occur in predicate positions a total of 30 times. The 10 remaining instances are those that were not repaired and cannot be clearly coded as fulfilling the syntactic projection of an intransitive or transitive verb.

Table 6: Placeholder frequencies of Nasal fillers by syntactic position.

	Nominal	Verbal
DEM.SP	1	-
DEM.ADDR	55	22
DEM DIST	14	4
Interrogative	6	4
Total: 106 =	76	30

about what was done with the fish he had seen earlier, using PV- and applicative-marked *di-ani-kun* ‘PV-DEM.ADDR-CAUS/BEN’ to fulfill the syntactic projection of a transitive verb describing ‘putting them in the freezer’. In such a case, the valency-increasing suffix *-kun* allows for the otherwise intransitive placeholder to take a Secondary Argument.

- (37) 1 E: *dipo iwo nu?*
 where fish earlier
 ‘where’s the fish from earlier?’
- 2 → *kan di:-ani-kun tian di kulkas kan?*
 TAG PV-DEM.ADDR-CAUS/BEN 3PL in freezer TAG
 ‘right, they already whatchamacallit-ed (=put) it in the freezer,
 right?’
- (BJM02-014-01, 00:18:35–00:18:39, Speaker: Edwar)

If a Secondary Argument is implicit or only stated in the repair, the placeholder need not have the applicative, as in (38) for AV and (39) for PV where the repair contains a P argument which is unrealized in the clause with the placeholder.⁶ In (38), Edwar is describing how he would use a kind of net, using the AV-marked placeholder *ng-ani* ‘AV-DEM.ADDR’ to fill the syntactic position of a verb meaning ‘pulling it in the shallows’. In (39), Gadis uses the distal demonstrative *udi* as a placeholder while she determines how to describe what Ros’s child is doing. Both cases are clearly verbal placeholders, the former because of the verbal marking and the latter because of the immediately preceding aspectual marker *masih* ‘still’.

⁶Although voice changes in the repaired clause, the transitivity does not – the placeholder is occupying the syntactic position of a verb indicating ‘pulling a net’, an action which certainly requires two arguments, as implied by the following PV transitive verb *takhik* ‘pull’.

- (38) 1 E: *amo kito pandai: ng-uyun ni agi*,
 if 1PL.INCL able AV-control DEM.SP again
 ‘if we can control it again,’
 2 → *nyak **ng-ani***,
 1SG AV-DEM.ADDR
 ‘I’m going to whatchamacallit,’
 3 *takhik di khampusan*
 PV.pull at shallows
 ‘pull (it) in the shallows.’
 (BJM02-014-01, 00:20:10–00:20:15, Speaker: Edwar)

- (39) 1 G: *kung*
 not.yet
 ‘not yet.’
 2 → *masih u:di*,
 still DEM.DIST
 ‘(he’s) still whatchamacallit,’
 3 *masih jeujo-kun=nyo*
 still PV.like.that-CAUS/BEN=3SG
 ‘still do them like that (=holding his hands up).’
 4 *kung ber-arti pedum na*
 not.yet MID-meaning sleep well
 ‘that means that he hasn’t fallen asleep yet.’
 (BJM02-050-03, 00:04:35–00:04:38, Speaker: Gadis)

6 Fillers in the larger corpus

A cursory look at our larger corpus of 24 transcribed conversations (~30 hours of speech) largely reflects what we have seen above. A superficial analysis of this set provides a broad overview of the range and various functions of each of the function words without respect to their polyfunctional uses, whether hesitators, placeholders, or within their typical demonstrative/interrogative contexts.

Frequency counts for each function word with various affixes are shown in Table 7. The nasal prefix is excluded from the discussion here since a large portion of forms affixed with *N-* must be attributed to polysemy/homophony – *ng-ajo* ‘AV-DEM.SP’, *ng-ani* ‘AV-DEM.SP’, and *ng-udi* ‘AV-DEM.SP’ with the emphatic demonstratives and *ng-api* ‘av-what’ with the interrogative *ngapi* ‘why’. When

not used in their bare form, Nasal placeholders are most frequently found occurring with verbal affixes. The most common (excluding the AV nasal prefix) is the applicative suffix *-kun* ‘CAUS/BEN’. Other typically verbal affixes such as *di-* ‘PV’, *te-* ‘NVOL’, and *be-* ‘MID’ are also found on Nasal placeholders, although these are rare in our corpus. The same can also be said of the nominalizing affixes *peN-*, *ke-*, and *-an*.

Table 7: Morphological marking on Nasal placeholders in the larger corpus.

	Verbal				Nominal		
	<i>di-</i>	<i>te-</i>	<i>be-</i>	<i>-kun</i>	<i>peN-</i>	<i>ke-</i>	<i>-an</i>
	PV	PASS	MID	APPL	AGENT	NOM	NOM
DEM.SP	1	-	-	1	-	-	-
DEM.ADDR	11	3	3	55	1	2	15
DEM.DIST	13	2	4	57	1	-	10
Interrogative	5	1	3	22	-	-	4

Frequency counts for the fillers occurring with each of the proclitics and enclitics in our corpus are shown in Table 8. The use of *=nyo* ‘3SG’ far exceeds that of the other clitics, unsurprisingly so given its polysemous nature as a possessive with nouns, A and P indexes in transitive verbs, the preference for third-person over first- and second-person referents in discourse, and the fact that *=nyo* ‘3SG’ in Nasal additionally operates as a marker of definiteness among other functions (see the discussion in §2.1). Because of this polysemy between verbal and nominal functions, *=nyo* is marked as ‘Ambiguous’ in the table.

Table 8: Cliticization on Nasal placeholders in the larger corpus.

	Verbal		Nominal		Ambiguous
	<i>ku=</i>	<i>mu=</i>	<i>=ku</i>	<i>=mu</i>	<i>=nyo</i>
	‘1SG’	‘2SG’	‘1SG’	‘2SG’	‘3SG’
DEM.SP	-	-	-	-	5
DEM.ADDR	3	-	1	-	44
DEM.DIST	1	-	1	1	57
Interrogative	-	-	1	1	95

While all of the most common affixes and clitics used with placeholders have been discussed above, those not already discussed but shown in Table 8 (e.g. *be-* or *peN-*) do not play a major role in the morphosyntax. Apart from clearly indicating placeholder rather than hesitator or demonstrative/interrogative uses of the function words discussed here, these affixes give little insight into the operation of fillers.

7 Some implications

The investigation of fillers in a corpus of six conversations has allowed us to see that, while all four function words investigated have the ability to take on hesitator and placeholder functions, the demonstrative and interrogative pronouns have complementary tendencies. Demonstrative fillers favor the placeholder functions, while the interrogative pronoun favors hesitator functions. Hayashi & Yoon (2006) have demonstrated how demonstratives have come to function as placeholders through their focusing/pointing function and as hesitators through a separate process of pragmaticization. Nasal provides additional support for their analysis of demonstrative fillers that serve as both placeholders and hesitators.

The person-oriented and tripartite division of demonstratives makes Nasal stand out against neighboring Malay varieties. Because Nasal lacks a dedicated filler (cf. Besemah *anu*, McDonnell & Billings 2025 [this volume]), demonstratives play a larger functional role in dealing with word-formulation trouble. Furthermore, Malay varieties most commonly have a two-way proximal-distal division in demonstratives, and when employed as a filler, it is the proximal that is most frequent (see McDonnell & Billings 2025 [this volume]; Wouk 2005). However, in Nasal the near-addressee demonstrative is by far the most frequent filler. This is perhaps due to the fact that the pragmatic role of placeholders (the most common role of demonstrative fillers in Nasal) is to signal contextual knowledge purported to be shared with the addressee or to invite the addressee in repair and word-search functions (see the discussion of participant access in Hayashi & Yoon 2006).

The origin of the use of the interrogative pronoun as a placeholder and as a solitary hesitator is not as clear as with the demonstratives, but parallels can be found elsewhere both within the Austronesian family (see Blake 2020 for Western Subanon, Tanangkingsing 2022 for Cebuano, Nagaya 2022 for Tagalog) and outside the family (see Buskunbaeva 2021 for Bashkir, Klyachko 2022 for Evenki). Given, however, that interrogatives, like Nasal *api*, frequently participate in the

development of indefinite pronouns (Haspelmath 2001), the referentiality of the indefinite pronoun (similar to that of a demonstrative as described above) could have given rise to its use in a placeholder function. It is possible that the placeholder uses of *api* derived from this use as an indefinite pronoun and its hesitator use from self-addressed questions, but further study, both for Nasal and cross-linguistically, is needed to understand the development of fillers from interrogative pronouns.

Placeholders are designed to fulfill the function of the word that a speaker intends to say, and they often do so by mirroring not only the syntactic position but the morphology of the word that is delayed. Fox (2010: 6) states this position clearly: “In languages that mark their placeholder fillers with complex morphology, that morphology may be part of the adequacy of the promise for fulfillment.” However, as Podlesskaya (2010) shows in her typological survey of fillers, placeholders need not precisely mirror the item they stand in for. In fact, given the hesitator nature of placeholders, it should not surprise us that affixation and cliticization differ between a placeholder and its target. Furthermore, such ‘fulfillment’ need not even take place. In opposition to Hayashi & Yoon’s (2010: 42) remark that “Frequently (but not invariably), a placeholder demonstrative is subsequently replaced”, repair in Nasal can scarcely be called frequent, occurring in less than two thirds of placeholder instances in our corpus. Despite this, Nasal speakers proceed without any apparent issue (see examples (26) and (34)–(37)).

As is the case elsewhere in Austronesia, we have seen how the lack of explicit arguments and explicit marking based on word class in combination with the variable word orders leads to indeterminacy in analysis of fillers as placeholders or hesitators. However, that the division between typical interrogative and demonstrative pronoun functions and their filler functions is not as clear-cut as those discussed by Hayashi & Yoon (2010) demonstrates both the great variety of fillers and the lack of depth of our knowledge about how fillers operate cross-linguistically. We suspect that with other languages of Austronesia (as, for example, is also the case with South Barisan Malay, McDonnell & Billings 2025 [this volume]) and beyond (as in Negidal, Pakendorf 2025 [this volume]), such pronouns may frequently vary in their usage and clear divisions between uses of hesitator and placeholder functions may not exist.

8 Conclusion

In this chapter, we described the various ways hesitator and placeholder fillers are utilized in everyday conversations in Nasal. We demonstrated how demonstratives and the interrogative *api* ‘what’ fulfill these functions. While there are other phrasal and non-lexical means of accomplishing the same two functions, these polyfunctional lexical items are frequently encountered in interaction as a means for dealing with trouble in word-formulation.

Appendix A Coding schema for conversational corpus

ID	Description	Levels
Form	What is the full morphological form of the filler?	e.g. <i>ani=nyo</i>
Root	What is the root of the filler?	e.g. <i>ani</i>
Function	What is the function of the filler?	placeholder, hesitator, indeterminate
IU Position	What position within the Intonation Unit (IU) does the filler occupy?	initial, medial, final, sole
Disfluency	Are there any disfluencies immediately preceding or following the filler?	truncation, sound stretch, pause, none
Syntactic position	What position in the clause does the filler occupy?	predicate, argument, modifier, PP object, adjunct
Repair	Is the placeholder repaired?	yes, no, NA
Recycling	Is any part of the phrase recycled?	yes, no, NA
Repair location (turn)	How many turns after the placeholder is used is it repaired?	0-5, NA
Repair location (IU)	How many IUs after the placeholder is used is it repaired?	0-9, NA
Repair initiator	Who initiates repair of the placeholder?	self, other, NA
Repairer	Who repairs the placeholder?	self, other, NA

Abbreviations

ADDR	Near-addressee	MID	Middle
AV	A-Voice	NEG	Negative
BEN	Benefactive	NVOL	Non-volitional
CAUS	Causative	PV	P-Voice
DEM	Demonstrative	RDP	Reduplication
DIST	Distal	REL	Relativizer
EMPH	Emphatic	SP	Near-speaker
LOC	Locative		

Acknowledgements

We would like to thank two anonymous reviewers and the editors for valuable feedback on this chapter. We would also like to acknowledge Johan Safri and Wawan Sahrozi, who have provided their expertise in the transcription, translation, and analysis of the Nasal conversations, as well as the Nasal people who have participated in the conversations in the corpus. We would like to thank our research counterpart in Indonesia, Yanti (Atma Jaya Catholic University of Indonesia), and the National Research and Innovation Agency (BRIN) for allowing us to conduct research on Nasal. This research is based upon work supported by the National Science Foundation under Grant BCS-1911641. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Anderbeck, Karl & Herdian Aprilani. 2013. *The improbable language: Survey report on the Nasal language of Bengkulu, Sumatra*. Dallas: SIL International.
- Billings, Blaine & Bradley McDonnell. 2024. Sumatran. *Oceanic Linguistics* 63(1). 112–174. DOI: 10.1353/ol.2024.a928205.
- Blake, A. L. 2020. Delays without fillers: The example of Western Subanon. *University of Hawai'i at Mānoa Working Papers in Linguistics* 51(1). 1–25. <http://hdl.handle.net/10125/73275>.
- Blust, Robert. 2013. *The Austronesian languages* (Pacific Linguistics Series C 602). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

- Blust, Robert, Stephen Trussel & Alexander D. Smith. 2023. *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2) [dataset]*. DOI: 10.5281/zenodo.7741197.
- Buskunbaeva, Liliya A. 2021. The functioning of the verbal hesitation marker *Ni/Nei* 'whatchamacallit' in oral monologues of Bashkirs: A case study of Bashkir dialectical texts. *Oriental Studies* 14. 172–185.
- Dimock, Laura. 2010. Fillers and placeholders in Nahavaq. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 119–138. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.06dim.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane Anne Edwards & Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale: Lawrence Erlbaum Associates, Publishers.
- Enfield, N. J. 2003. The definition of *what-d'you-call-it*: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Ewing, Michael C. 2005. *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Amsterdam; Philadelphia: John Benjamins.
- Ewing, Michael C. 2019. The predicate as a locus of grammar and interaction in colloquial Indonesian. *Studies in Language* 43(2). 402–443. DOI: 10.1075/sl.17062.ewi.
- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders*, 1–10. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.01fox.
- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Gil, David. 2003. Word order without syntactic categories: How Riau Indonesian does it. In Andrew Carnie, Heidi Harley & Sheila Dooley (eds.), *Verb first: On the syntax of verb-initial languages*, vol. 73 (Linguistics Today), 243–263. Amsterdam: John Benjamins.
- Haspelmath, Martin. 2001. *Indefinite pronouns* (Oxford studies in typology and linguistic theory). Oxford: Oxford University Press.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of

- word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 33–66. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.03hay.
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90. 927–960. DOI: 10.1353/lan.2014.0105.
- Kasim, Yuslina, Marjusman Maksan, Syamsir Arifin, Zailoet & Muhamad Yamin. 1987. *Pemetaan bahasa daerah di sumatra barat dan Bengkulu*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Kluge, Angela. 2017. *A grammar of Papuan Malay* (Studies in Diversity Linguistics 11). Berlin: Language Science Press. DOI: 10.5281/zenodo.376415.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hödl & Thomas Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- McDonnell, Bradley. 2017. *Documentation of Nasal: An overlook Malayo-Polynesian isolate of southwest Sumatra*. London: Endangered Languages Archive. <http://hdl.handle.net/2196/00-0000-0000-0010-798B-E>.
- McDonnell, Bradley. to appear. Documenting multilingualism in Southwest Sumatra. In Maya Ravindranath Abtahian, Abigail C. Cohn, Thomas J. Connors, Joseph J. Errington & Yanti (eds.), *Indonesian languages and linguistics: State of the field*. Ithaca: SEAP Publications, Cornell University Press.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- McDonnell, Bradley, Blaine Billings & Jacoob Hakim. 2019. *Documentation of the multilingual linguistic practices of the Nasal speech community*. Sydney: Collection BJM02 at catalog.paradisec.org.au. DOI: 10.26278/5f46870d43f29.
- McDonnell, Bradley & Christina L. Truong. 2024a. Applicative constructions in languages of Western Indonesia. In Fernando Zúñiga & Denis Creissels (eds.), *Applicative constructions in the world's languages*, 971–1004. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110730951-028.
- McDonnell, Bradley & Christina L. Truong. 2024b. Non-Malayic languages of Sumatra and the Barrier Islands. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*,

- 409–430. Oxford: Oxford University Press. DOI: [10.1093/oso/9780198807353.003.0028](https://doi.org/10.1093/oso/9780198807353.003.0028).
- McDonnell, Bradley, Jiang Wu, Timothy McKinnon & Alexander Adelaar. 2024. The Malayic languages. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*, 431–454. Oxford: Oxford University Press. DOI: [10.1093/oso/9780198807353.003.0029](https://doi.org/10.1093/oso/9780198807353.003.0029).
- Mosel, Ulrike. 2023. Word classes in Austronesian languages. In Eva van Lier (ed.), *The Oxford handbook of word classes* (Oxford Handbooks in Linguistics), 771–787. New York: Oxford University Press.
- Næss, Åshild, Anna Margetts & Yvonne Treis. 2020. *Introduction: Demonstratives in discourse*. Åshild Næss, Anna Margetts & Yvonne Treis (eds.) (Topics at the Grammar-Discourse Interface 6). Berlin: Language Science Press. 1–20. DOI: [10.5281/zenodo.4054814](https://doi.org/10.5281/zenodo.4054814).
- Nagaya, Naonori. 2022. Beyond questions: Non-interrogative uses of *ano* ‘what’ in Tagalog. *Journal of Pragmatics* 190. 91–109. DOI: [10.1016/j.pragma.2022.01.007](https://doi.org/10.1016/j.pragma.2022.01.007).
- Németh, Zsuzsanna. 2012. Recycling and replacement repairs as self-initiated same-turn self-repair strategies in Hungarian. *Journal of Pragmatics* 44. 2022–2034.
- Pakendorf, Brigitte. 2025. A corpus-based investigation of fillers in Negidal (Northern Tungusic). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitations and placeholders*, 75–129. Berlin: Language Science Press. DOI: [10.5281/zenodo.15697579](https://doi.org/10.5281/zenodo.15697579).
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: [10.1075/tsl.93.02pod](https://doi.org/10.1075/tsl.93.02pod).
- Stokhof, Wim A. L. 1987. *Holle lists: Vocabularies in languages of Indonesia/2, Southern Sumatra*, vol. 10 (Pacific Linguistics Series D 75). Canberra: The Australian National University.
- Streeck, Jürgen. 1996. A little Illokano grammar as it appears in interaction. *Journal of Pragmatics* 26. 189–213.
- Tanangkingsing, Michael. 2022. Pragmatic functions of versatile *unsa* ‘what’ in Cebuano: From interrogative pronoun to placeholder to stance marker. *Journal of Pragmatics* 193. 59–75.
- Truong, Christina L. & Bradley McDonnell. 2022. Neglected functions of Western Indonesian applicatives. In Sara Pacchiarotti & Fernando Zúñiga (eds.), *Applicative morphology: Neglected syntactic and non-syntactic functions*, 405–436. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110777949](https://doi.org/10.1515/9783110777949).

- Williams, Nicholas. 2010. Toward a linguistic anthropological account of deixis in interaction: ini and itu in Indonesian conversation. *Colorado Research in Linguistics* 22. DOI: 10.25810/0GHQ-WK77.
- Woollams, Geoff. 1996. *A grammar of Karo Batak, Sumatra* (Pacific Linguistics Series C 130). Canberra: The Australian Nationa University.
- Wouk, Fay. 2005. The syntax of repair in Indonesian. *Discourse Studies* 7(2). 237–258.

Chapter 7

Form and frequency of Kalamang fillers

Eline Visser

Uppsala University

This contribution focuses on the form and frequency of fillers in Kalamang (Greater West Bomberai, Papuan), based on a 7416-word corpus. It reveals more forms for both placeholders and hesitations than previously described for this language. The main placeholder is *neba* ‘what’, which can replace nouns and verbs. It is used with a much higher ratio than previously calculated (9.84 per 1000 words instead of 6), which is also higher than placeholder use reported in other languages, such as Russian and Mandarin (5 and 6.68 per 1000 words, respectively). The most common hesitative is *nain* ‘like’, followed by *a* ‘uh’. Coding of the phonetic form of all hesitations revealed many different pronunciations, however, with clear inter-speaker preferences for [nain] or [nain:] and [a:], respectively. This data can help in making informed choices in the development of a standardised spelling. Finally, the data reveal big differences between speakers in frequency of use of fillers (ranging from 0 to 92.36 per 1000 words), stressing the need for speaker-balanced corpora.

keywords: Papuan, placeholder, hesitator, inter-speaker variation

1 Introduction

Kalamang (kgv, kara1499) is a Greater West Bomberai language (Usher & Schapper 2022) spoken on the biggest of the Karas Islands in Fakfak Regency in the province of West Papua in Indonesia. The map in Figure 1 shows the Karas Islands.

Kalamang is spoken by around 130 people in two villages on the biggest of the Karas Islands: Mas and Antalisa. Kalamang is under pressure from the local lingua franca, a variant of Papuan Malay (PM), and is not currently spoken by people born after 1990. The texts in this study are all conversations that were





Figure 1: The location of Karas, with the names of the six villages on the Karas Islands in Kalamang (*italics*) and Indonesian. Kalamang is spoken on the biggest island.

recorded between 2017 and 2019 as part of the author's PhD project, which resulted in a comprehensive grammar of Kalamang (published later as Visser 2022). All Kalamang linguistic and cultural data have been deposited in Lund University's Humanities Lab Archive (Visser 2020).

The rest of this paper is structured as follows. In the rest of the introduction, I describe the different Kalamang filler forms (§1.1) and the data and methods used in this study (§1.2). Then, in §2, I discuss the findings from the selected corpus study for placeholders, and in §3 I do the same thing for hesitatives. §4 presents the conclusions.

1.1 Description of Kalamang filler forms

In this section, I will introduce the forms that are further discussed in §2 and §3. The examples in this section are taken from recordings in the full Kalamang corpus (Visser 2020). The current study, reported from §2 onwards, focuses on a subset of the corpus, introduced in §1.2.

A placeholder is defined as a lexical form that replaces a target item (Hayashi & Yoon 2006: 490, Podlesskaya 2010: 11). A hesitative is defined as an overt marker of hesitation that is not syntactically integrated (Hayashi & Yoon 2010: 35), here restricted to conventionalized markers of hesitation.

There is no overlap in forms between placeholders and hesitatives in Kalamang: the forms that are used as placeholders cannot be used as hesitatives and vice versa.

1.1.1 *Neba* 'what'

The placeholder *neba* 'what' is homonymous with the nominal question word *neba* 'what'. The question word is likely to be the source for the placeholder (Podlesskaya 2010: 12). Placeholder *neba* 'what' targets nouns and verbs and occurs in the same slot as its target. Although its likely origin is nominal, both noun phrase and predicate morphology can attach directly to the root. It is not found to replace other word classes like adverbs. Most¹ NP and predicate morphology that is known for Kalamang is attested on placeholder *neba*. All morphology attested on placeholder *neba* is found on nouns and verbs. *Neba* typically carries all morphology that the replaced noun or verb would have carried, morphosyntactically behaving as the target. The following examples show some of the uses of this placeholder.

¹With the exception of some rarer forms, which are probably absent because of their low frequency rather than because of some rule that prohibits them from occurring on the placeholder.

(1) and (2) illustrate *neba* with nominal targets. In (1), *neba* replaces a location and is inflected with allative/ablative *=ka*. In (2), *neba* replaces a noun and is inflected with the third-person possessive marker *-un* and object marker *=at*. In both cases, the target is retrieved immediately after, and we see the same markers on the target.

- (1) *yuol me in bara kai-rep=ta neba=ka Sowir=ka*
day TOP 1PL.EXCL descend firewood-get=NFIN what=LAT Sowir=LAT
'One day we went down to get firewood at whatsit, at Sowir.' [conv11_53]
- (2) *pi neba-un=at pes-un=at dikolko*
1PL.INCL what-3POSS=OBJ peel-3POSS=OBJ remove
'We remove its whatsit, its peel.' [narr14_17]

(3) and (4) illustrate *neba* with verbal targets. In (3), *neba* is negated with negator *=nin*, as is the target *namakin* 'to fear'. In (4) *neba* carries irrationalis marker *=et*, while the target is not retrieved.

- (3) *per=at kan ma neba=nin to namakin=nin*
water=OBJ you.know 3SG what=NEG right fear=NEG
'Water, you know, it doesn't whatsit, right, doesn't fear.' [conv13_180]
- (4) *sontum neba=et mu jaga=te*
people what=IRR 3PL keep.watch=NFIN
'[If] people are whatsitting they watch out.' [conv9_382]

Neba may also be used for entities that the speaker does not know the name of. It is in that case still a placeholder, but there is no intention to retrieve the target. Its use is very similar to English *what-d'you-call-it* as described in Enfield 2003. In (5), the speaker describes a picture of a wooden toy construction made with Tinkertoy, which the speaker is not familiar with. He describes two of the parts with *neba*, leaving it up to the addressee to identify the correct referent. The addressee, whose task it is to find a picture matching the description, can deduce which parts the speaker intends from other information (such as the numeral *eir* 'two' and the locational noun *raor* 'middle').

- (5) *wa me neba-un eir kareta-un iriskap neba-un raor=ko*
PROX TOP WHAT-3POSS two cart-3POSS white what-3POSS middle=LOC
'This has two whatsits, a white cart, a whatsit in the middle.' [stim39_42]

1.1.2 *Don* ‘thing’

Don ‘thing’ is a generic noun. It is used as a nominal placeholder when the speaker wants to avoid expressing the target, and hence the placeholder *don* is used without a target. *Don* ‘thing’ is frequently used as something like a code word (when the regular term must be avoided), and to express disdain. Using the terminology of Seraku (2024), *don* is a placeholder, the use of which is motivated by the speakers’ preferences, not by their abilities. Placeholders that are used in avoidance strategies are also found in Chinese (Cheung 2015), Lao (Enfield 2003), Komnzo (Döhler 2025 [this volume]) and Nasal (Billings & McDonnell 2025 [this volume]).

(6) and (7) illustrate the generic use of *don*. In (6) it is incorporated in the verb *paning* ‘ask’, while in (7) it is modified by a proximal demonstrative.

- (6) *ka me don-paning=sawe*
2SG TOP thing-ask=too
'You ask for too much.' [conv10_224]
- (7) *mu don yua=at napasang*
3PL thing PROX-OBJ hang.up
'They hang up this stuff.' [conv10_3]

Examples of code words are given in (8). These are used when the regular version is inappropriate: for example, when begging someone else for these goods (which may be scarce), or when communicating with someone in your household about the lack of these items in front of a guest.²

- (8) a. *don pen~pen*
thing sweet~sweet
'sugar'
- b. *don iriskap*
thing white
'rice, sugar'
- c. *don yuolyuol*
thing shine
'lamp' [dictionary]

²Code words may also be made without *don* ‘thing’. Another code word for sugar or rice is *muap iriskap*, lit. ‘white food’, and another code word for *pitis* ‘money’ is *lolok* ‘leaf’.

In (9), taken from a traditional story, instead of saying *sor* ‘fish’, the narrator uses derogatory *don* to express his disdain towards the fact that a crow has eaten rotten fish.³

- (9) *o ka don yuwa=at=a na=tauna sehingga don mun=ten*
EMPH 2SG thing PROX=OBJ=FOC consume=so so_that thing rotten=AT
wandi=et ka bisa na=ta
like_this=IRR 2SG can eat=NFIN
'Oh, you eat this stuff, so that [this means] you can eat rotten stuff like
this.' [narr39_130]

I consider the uses of *don* in (8) and (9) to be placeholders.

In contrast to placeholder *neba* ‘what’, *don* ‘thing’ is not typically used when the speaker has trouble retrieving the target. For example, *don* is very seldomly used in combination with hesitatives, and as stated above, the target is not expressed with this placeholder. One possible counterexample is (10). Here, *don* is followed by hesitative *a* ‘uh’. It is not clear whether *don* is a placeholder for an unretrieved target (perhaps a modifier to rice), or for the verb *kuar* ‘to cook’ (which would be strange, since it otherwise is a placeholder for nouns). An alternative analysis is that *don* in (10) is a hesitative. Since there are no other examples like this one, I will continue analyzing *don* ‘thing’ as a nominal placeholder used when the speaker is not willing to verbalize the target.⁴

- (10) *in me pasa me don a kuar me*
1PL.EXCL TOP rice TOP thing uh cook TOP
'As for us, considering rice, whatsit/like, cooking...' [conv13_102–103]

1.1.3 *Apa* ‘what’

Apa ‘what’ is a question word and filler borrowed from Indonesian. There is only one clear example, given in (11), where it can be seen that the *apa* does not carry the same marking as the target.

- (11) *koi nain (0.6) apa nain olun=kin=at bisa to*
then like what like leaf=POSS=OBJ can right
'Then like... whatsit, [we] can [talk about] like, leave's [medicines],
right?' [conv20_158]

³This seems to be an uncommon use of *don* ‘thing’.

⁴Contra Seraku (2024: 16), who opens up for the possibility that *don* can also be used when the speaker is unable to retrieve the target, based on communication with me. In hindsight, I see no indication that *don* is ever used in these inability contexts.

From the one corpus example we cannot tell whether this filler should rather be counted as a placeholder or as a hesitative. However, from my experience with Kalamang, I know it to be used as a placeholder, so for now I count it as such. Klyachko (2025 [this volume]) and Ventayol-Boada (2025 [this volume]) also report on borrowed fillers in Tungusic languages and Kolyma Yukaghir, respectively.

1.1.4 *Puraman* ‘how many’

The question word *puraman* ‘how many’ is used as a placeholder for numbers and as an indefinite numeral. It does not occur in the sample used in the quantitative part of this study (§2), but I will briefly illustrate it with examples from the full corpus here.

As an indefinite numeral, it refers to a large-ish number, the exact amount of which the speaker does not know or does not feel the need to convey. In (12), *puraman* is used to convey that there were many children, but the speaker doesn’t know the exact amount.

- (12) *kewe tumtum puraman=bon neko*
 house children how_many=COM inside
 ‘A house with I-don’t-know-how-many children inside.’ [stim43_95–96]

(13) is a genuine placeholder, which is used to fill the slot of the numeral until the target ('three') is found. When *puraman* ‘how many’ is used as a placeholder, it carries the same inflection as the quantifier it replaces, and occurs in the same slot. Note, however, that although the numeral in (13) is suffixed to the pronoun *mu*, *puraman* is not.⁵

- (13) *afukarun nasuarik, mu puraman? Munggaruok. Munggaruok*
 afukat-un nasuarik mu **puraman** mu-karuok mu-karuok
 avocado-3POSS scatter 3PL how_many 3PL-three 3PL-three
mat rupte kajie
 mat rup=te kajie
 3SG.OBJ help=NFIN pick

‘His avocados scatter, how many of them? The three of them. The three of them help him pick [the avocados up].’ [stim29_23]

⁵Otherwise, /p/ would lenite to [w]: *muwuraman* (Visser 2022: 96).

1.1.5 *Nain* ‘like’

Finally, the word *nain* ‘like’ is used as a hesitative. In (14), the speaker has a false start, starts the sentence, uses the hesitative, and then finally utters the whole sentence.

- (14) *ah, ka-pi he nain, pi he bo lalaret*
ah ka-pi he **nain** pi he bo lalat=et
INT 1PL.INCL IAM like 1PL.INCL IAM go die=IRR
'Ah... until we like, until we die.' [conv20_229]

Nain also has a literal use in comparative constructions, illustrated in (15). Comparative constructions can usually be recognized by the similitative enclitic =*kap*.⁶

- (15) *pi na-te nain dongdong-ten=kap telin*
1PL.INCL consume=IRR like chewy-AT=SIM stay
'If we eat it, it's like very chewy.' [conv13_125]

1.2 Data and methods

The data used in this study comes from a subset of the Kalamang corpus (Visser 2020), which was re-annotated with special attention paid to the phonemic form and gloss of fillers. The dataset consists of seven texts involving eight different speakers. There are three men (Kamarudin, Hair and Sabtu) and five women (Fajaria, Nurmia, Bini, Hawa and Samsia). All texts are conversations. Totaling 7416 words, the texts contain 61 hesitations and 77 placeholders. Disfluencies like lengthening, mumbling and repetition are classified as ‘other’ and discussed together with hesitations. Table 1 gives an overview of the seven texts. The tags in the table (as well as the tags in the examples in this chapter) are clickable and lead to the bundle page in the archive containing the relevant files. All files are openly accessible.

All these texts were initially transcribed, translated and glossed in the period between 2017 and 2020. For this study, I re-listened to all recordings, paying special attention to fillers. This was necessary because some fillers were not transcribed in the initial round at all, while some instances of placeholders were

⁶Comparative constructions can also be made without similitative =*kap*. They can then be distinguished from hesitative *nain* by the elements present in the clause. In a comparative construction there is a standard, a parameter and a comparee. If one of these are lacking, *nain* is analysed as a hesitative.

Table 1: The texts used in this study

Title	Tag	Speakers	Words	H ^a	P ^b	O ^c
Binkur mama and Bakri mama talk current affairs	conv9	Fajaria, Nurmia	2170	3	17	6
A kitchen conversation between two grandmothers	conv12	Bini, Hawa	1514	0	10	1
Mohtar's father and Lamani's father discuss root medicines	conv20	Kamarudin, Hair	1256	45	19	16
A conversation about chestnuts	conv11	Fajaria, Nurmia	1074	5	20	8
Two grandfathers talk current affairs	conv14	Kamarudin, Sabtu	675	8	4	3
Netfishing 3	conv3	Fajaria, Samsia	379	0	3	0
Netfishing 2	conv2	Fajaria, Samsia	348	0	4	0
	total		7416	61	77	34

^aHesitatives^bPlaceholders^cOther

analyzed as another word class. *Neba* is a question word when the speaker clearly invites the addressee to answer. It is a placeholder when the speaker does not, for example when the speaker is clearly looking for the target themselves, when the speaker retrieves the target, or when the speaker immediately continues their utterance, not wishing to retrieve the target. In unclear cases, the instance of *neba* is counted as a question word. *Don* is a noun when it is used in a generic sense. When it is not, it is a placeholder. In unclear cases, it is counted as a noun. Fillers were classified as placeholders when an educated guess could be made about which referent it stood in for.

Of two texts only a part was used, because they were very long: of conv9, the first 22 minutes (or 2170 words) were reanalyzed, and of conv12, the first 10 minutes (or 1514 words). Where fillers weren't transcribed yet, I added them to the transcription in ELAN. Furthermore, I added a tier for notes about the form and possible function of the filler. Further annotation was done in a spreadsheet,

where I also added the exact phonetic form, a confidence level (high/low), and for placeholders, whether or not the target was retrieved. All rows with a low confidence level ($n = 30$) were ignored in the final analysis.⁷ The data was further analysed in R. The script and the spreadsheet are available at OSF.⁸

2 Placeholders

2.1 Overview

The sample for the current study contains three different placeholders: *neba* ‘what’ (73 occurrences), *don* ‘thing’ (3 occurrences) and *apa* ‘what’ (1 occurrence). *Puraman* ‘how many’, which was found as a placeholder for quantifiers in Visser (2022) and was described in §1.1.4, was not encountered in the sample for the current study. The total number of placeholders is 77.

The frequency of the placeholders encountered in this study are given in Figure 2.

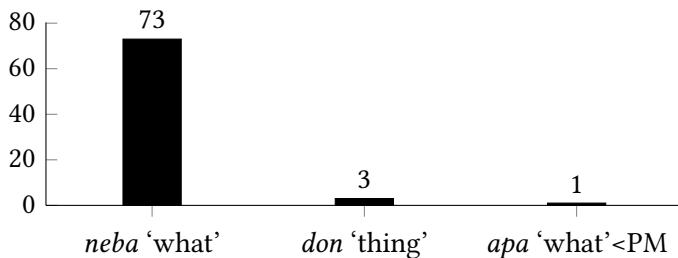


Figure 2: Frequency of placeholders.

2.2 Frequency

Figure 3 shows that all six speakers that use placeholders use *neba* ‘what’. The frequencies of the other placeholders are so low that one cannot meaningfully generalize.

Placeholder *neba* was previously reported to occur with a frequency of 6 per 1000 words in the full Kalamang corpus (Visser 2022: 431).⁹ This frequency is

⁷Table 1 only shows the counts of fillers with a high confidence level.

⁸https://osf.io/m9v4r/?view_only=ae5888aa7391484f9b61811d8c5bec6f

⁹That means all narratives and conversations, including those prompted with some kind of stimulus material, but excluding elicitation sessions (which were not recorded at all). There

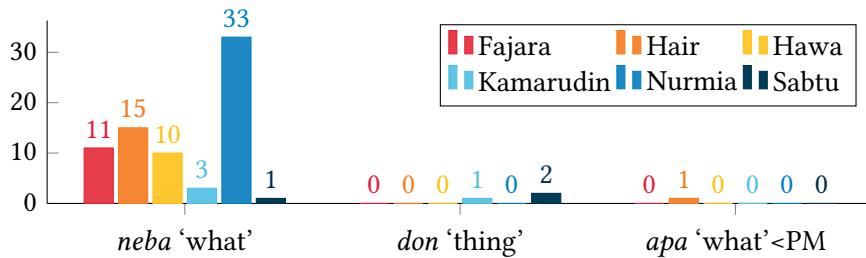


Figure 3: Usage of placeholders per speaker.

higher in the current sample: $73/7416 \times 1000 = 9.84$. If we also take into account the other placeholders, this ratio rises to 10.38 per 1000 words.

As for target retrieval, of the 73 instances of *neba* ‘what’, 25 times the target was retrieved (34.2%) and 42 it was not (57.5%). In the other 6 cases, it was unclear whether the target was retrieved or not. In the only example with *apa* ‘what’, the target was retrieved. In the three examples with *don* ‘thing’, no targets were retrieved. This is as expected, since *don* is used to avoid expressing the target.

There are large inter-speaker differences in placeholder frequency, as presented in Table 2. While Hair uses more than 28 placeholders per 1000 words, Kamarudin does not even use four, and two speakers, Samsia and Bini, use no placeholders at all. See also Ponsonnet (2025 [this volume]) for differences between speakers, there described as differences in “disfluency styles”, and Paken-dorf (2025 [this volume]), Ventayol-Boada (2025 [this volume]) and McDonnell & Billings (2025 [this volume]) for more observations on inter-speaker differences.

2.3 Discussion

This study shows that *neba* ‘what’ is very much the preferred placeholder. This mirrors findings from e.g. English, where variants of one placeholder (*thing(s)*, *thingy* and *thingie*) account for 75 to 87% of all placeholder occurrences (Palacios Martínez & Núñez Pertejo 2015) in four different corpora. The current study also reveals one instance of a borrowed placeholder: *apa* ‘what’ from Papuan Malay. This word is also used as a placeholder in Papuan Malay (Kluge 2017: 316). Like *neba* ‘what’, it has its origin in a question word, but it seems less

are 104 texts of in total 69706 words, spoken by 25 speakers (Visser 2022: 21, 28). The list of texts can be found in Appendix C of Visser (2022).

Table 2: Placeholder ratio per 1000 words per speaker.

speaker	ratio	words	placeholders
Hair	28.42	563	16
Nurmia	12.80	2579	33
Hawa	10.74	931	10
Fajaria	10.06	1093	11
Sabtu	8.33	360	3
Kamarudin	3.95	1012	4
Samsia	0	246	0
Bini	0	563	0

morphosyntactically integrated in the Kalamang grammar than *neba*, as in the one example it does not carry the inflection its target does (see 11 above).¹⁰

The adjusted frequency of placeholder *neba* ‘what’ from 6 in Visser (2022) to 9.84 here is likely due to two reasons: genre and more careful coding. In the current sample, only conversations were used. These are less planned in nature than narratives, which could possibly have an effect on placeholder use.¹¹ In earlier annotations, many uses of *neba* as a placeholder were coded as instances of the question word. Taking into account also the three instances of *don* ‘thing’ and the instance of *apa* ‘what’, the placeholder ratio rises to 10.38. This is much higher than what was previously found in large well-studied languages. Russian has 5 per 1000 words in a corpus of informal elicited narratives (Podlesskaya & Kibrik 2006 as cited in Podlesskaya 2010), and Mandarin has 6.68 per 1000 words in a corpus of telephone conversations (Zhao & Jurafsky 2005). For English, frequencies are much lower still. Palacios Martínez & Núñez Pertejo (2015) find frequencies of 0.03 to 0.07 in four different conversational corpora. However,

¹⁰In the entire Kalamang corpus, there are also a few instances of *apa lagi*, which literally means ‘what else’, but also ‘especially’ in Indonesian and Papuan Malay. In the Kalamang corpus, it seems also to be used as a placeholder. There are no other clear examples of placeholder *apa* in the entire Kalamang corpus.

¹¹But see Pakendorf (2025 [this volume]) for an example of a supposedly excellent speaker who has a high frequency of fillers in fairy tales in Negidal. And in Mojenó Trinitario, traditional narratives show the highest rate of placeholders (Rose 2021), often followed by a delayed constituent. Because the content of traditional stories is known to the speaker and most of the addressees, the major challenge for the speaker is not their competence in remembering the content, but their performance in rendering the story, hence the many hesitations in looking for the best wording.

recent studies on lesser-known languages find higher ratios: the Evenki filler *anj* has a frequency of 12.6 per 1000 words (Klyachko 2022), and Pakendorf (2025 [this volume]) finds 11.6 fillers (this includes both placeholders and hesitators) in Negidal. Combined with the Kalamang findings, this stresses the need for investigating placeholder frequencies in a wide range of languages.

Finally, we see large differences between speakers, which might boil down to differences in topic. Two of the eight speakers in the sample don't use placeholders at all. Note that these speakers' words form a small proportion of the total number of words (Samsia 3.3%, Bini 7.6%). Samsia has no other recordings in the entire Kalamang corpus, but there are five other recordings with Bini (both narratives and conversations). A quick search reveals that she uses *neba* 'what' as a placeholder five times in these recordings. A refined annotation might reveal more instances. On the other end of the scale is Hair, who has a ratio of 28.42, more than twice as high as number two, Nurmia. Like Samsia and Bini, in the sample he is represented with relatively few words, which all come from the same text. Hair is represented with other texts in the entire Kalamang corpus, but they are all narratives. In his longest narrative, which counts 1499 words, he does not use *neba* 'what' as a placeholder at all. An explanation for the high ratio in the present sample is that in the text (conv20), this speaker has great difficulties in retrieving the different herbal medicines that he knows, while the other speaker keeps urging him to do the talking. It was the researcher who requested the speakers to talk about herbal medicines. The topics of the other texts in the sample were determined by the speakers themselves. If we remove Hair from the sample, we are left with 61 placeholders and 6784 words, resulting in a ratio of 8.99 per 1000 words; still higher than what has been found in other languages.

3 Hesitatives and other disfluencies

3.1 Overview of the forms

Besides conventionalized *nain* 'like' (which is by far the most common with 37 instances): the sample also shows a high number of the less conventionalized *a* 'uh'¹² (24 instances). Apart from these two hesitatives, there are three disfluency strategies for gaining time or marking problems with the planning and execution of speech: lengthening, mumbling and repetition. All hesitatives and other disfluencies are given with their frequencies in Figure 4. There are 61 hesitatives and 34 other disfluencies in the corpus.

¹²This spelling of the English hesitative *uh*, *er* or *um* is used by for example Merriam-Webster (<https://www.merriam-webster.com/dictionary/uh>).

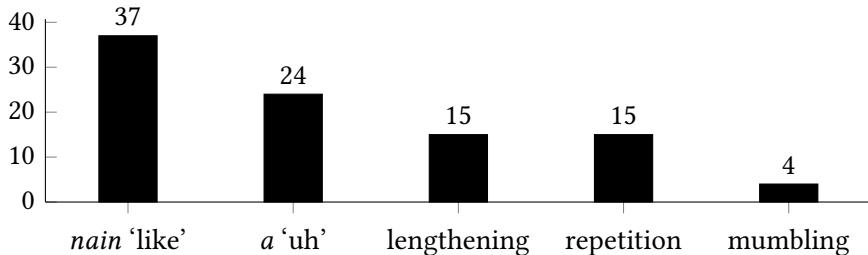


Figure 4: Frequency of hesitations and other disfluencies.

The sample shows that hesitations with a more or less conventionalized form (*nain* ‘like’ and *a* ‘uh’) are much preferred to other strategies that mark disfluency.

Nain ‘like’ is the most frequent hesitation, and it is often used together with a placeholder. It is also not uncommon to have two instances of *nain* in one utterance, like in (16).

- (16) *mu-nan nain opa nain neba-un me et kinkin=a saerak*
 3PL-too like ANA like what-3POSS TOP canoe small=FOC NEG.EXIST
leng wa me
 village PROX TOP
 ‘They too, like earlier, like whatsit, there are no small canoes in this
 village.’ [conv14_66]

Then in the next utterance, the speaker possibly says what he really wanted to say:

- (17) *emumur et kinkinun me saerak*
 women canoe small TOP NEG.EXIST
 ‘The women don’t have small canoes.’ [conv14_67]

Speakers can also start a turn with hesitative *nain* ‘like’, in the case of (18) followed by hesitative *a* ‘uh’. The pauses are indicated too, to show how much the speaker is struggling to formulate his thoughts.

- (18) *nain a (0.6) opa (1.2) sontum (0.8) gier-un=at ning sontum*
 like uh ANA people tooth-3POSS=OBJ ache people
mudamuda-ten gier-un=at taham=et
 young-AT tooth-3POSS=OBJ last=IRR
 ‘Like uh, before, people, whose teeth ache, young people who want to
 have lasting teeth...’ [conv20_193–196]

(19) illustrates *a* ‘uh’. It is followed by a brief pause, after which the speaker continues with some production errors (*se belum* ‘already not yet’) before he lands on what he wants to say (*pi tok muapnin* ‘we haven’t eaten yet’).

- (19) *tak ramandalin-i jien terus a (0.5) me me se belum*
 CLF.LEAF seven-OBJQNT get then uh DIST TOP already not_yet
pi tok muap=nin
 1PL.INCL still eat=NEG
 ‘[You] get seven leaves, then, uh, that’s when already, not yet, we haven’t
 eaten yet.’ [conv20_166]

(20) is an example of repetition, used in combination with the placeholder *neba* ‘what’.

- (20) *ge mena go ma ma ma neba*
 no otherwise condition 3SG 3SG 3SG what
 ‘No otherwise it, it, it is whatsit?’ [conv9_298]

The examples above show that hesitators and disfluencies may combine with pauses and other production difficulties, as well as placeholders (as in 16 and 20). (21) is a striking example that shows a combination of *nain* ‘like’, *a* ‘uh’ and *neba* ‘what’.

- (21) *terus nain nain sotnum gier-un ning a gier-un ning a*
 then like like people tooth-3POSS be_sick uh tooth-3POSS be_sick uh
neba ah met me nain
 what well DIST_OBJ TOP like
 ‘Then like, like, [when] people have a toothache, uh, a toothache, uh,
 whatsit, well that’s like...’ [conv20_41]

3.2 Frequency

Figure 5 shows that all types of disfluencies are used by all or most speakers that use them, except for mumbling, which is used by Hair only.

Nain ‘like’ has a ratio of 4.99 per 1000 words, while *a* ‘uh’ has a ratio of 3.24 (compare with 4.01 for *mm* and *uh* in Mandarin; Zhao & Jurafsky 2005).

There are large inter-speaker differences in the frequency of hesitative and other disfluency use, as presented in Table 3. Hair uses more than 92 hesitations per 1000 words, while the others score below 15. Two speakers, Samsia and Bini, use no hesitations at all.

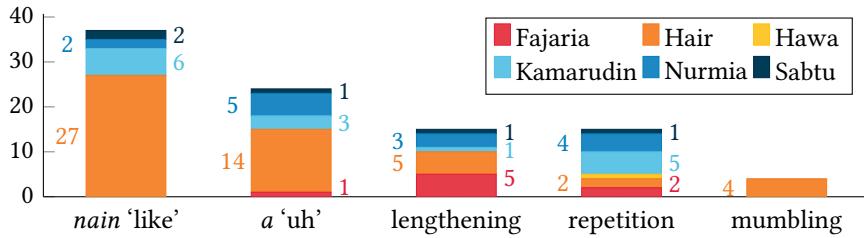


Figure 5: Usage of hesitations and other disfluency strategies per speaker.

Table 3: Hesitations and other disfluencies ratio per 1000 words per speaker.

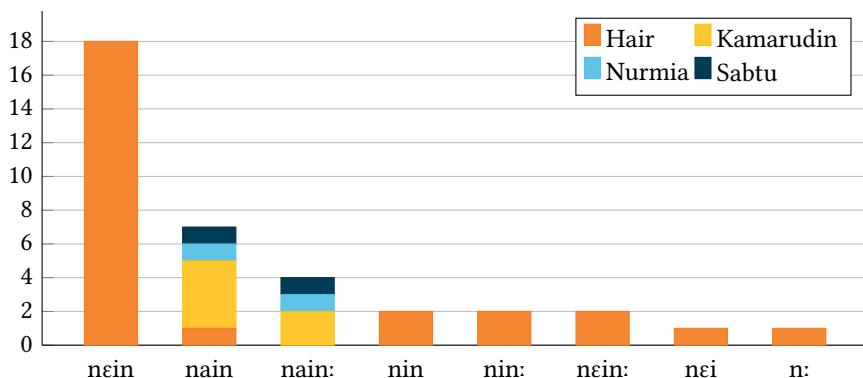
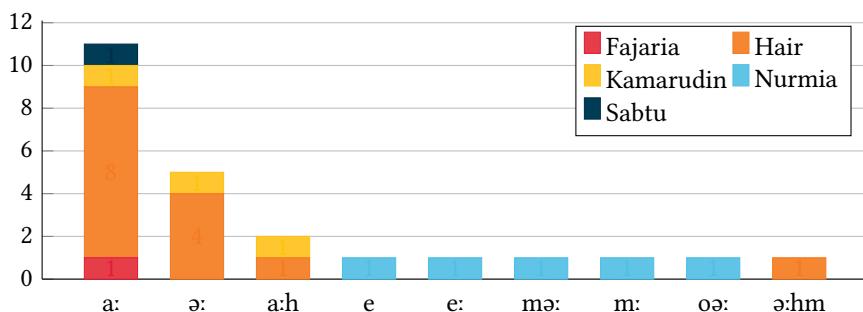
Speaker	Ratio	Words	Hesitations and other disfluencies
Hair	92.36	563	52
Kamarudin	14.82	1012	15
Sabtu	13.89	360	5
Fajaria	7.32	1093	8
Nurmia	5.43	2579	14
Hawa	1.07	931	1
Samsia	0	246	0
Bini	0	563	0

3.3 Pronunciation

There is quite some variation in the pronunciation of the hesitations *nain* and *a*. Figure 6 shows the variation for *nain* ‘like’, showing that the majority of the utterances is [nein] followed by [nain]. It also shows that while [nein] (and a few other variants) are exclusively used by Hair, while (lengthened) [nain] is used by all four speakers who use this hesitative.

Figure 7 shows the pronunciations attested for *a* ‘uh’. By far the most common pronunciation is [a:], and it is used by four of the five speakers who use this hesitative.¹³

¹³Kalamang also has an utterance divider *ah*. Its function seems to have something to do with structuring the discourse, akin to English conjunctions like ‘and then’ or ‘so’. The utterance divider is always pronounced as [a] (not lengthened), followed by a brief pause. It always occurs utterance-initially. An example can be found in the last part of (21).

Figure 6: Pronunciation variants of *nain* ‘like’.Figure 7: Pronunciation variants of *a* ‘uh’.

3.4 Discussion

This study has confirmed *nain* ‘like’ as the main hesitative. This word originally denotes similarity or attribution to a class, just like the hesitative *like* in English (Laserna et al. 2014) or *tipa* in Russian (Egorova 2018). The grammaticalization path from simile to quotative is well-attested across the world’s languages (Heine & Kuteva 2002: 274). From there, it is a small step into being pragmaticized into a word search marker or a hesitative (Vera Podlesskaya, p.c.). Note that Kalamang *nain* ‘like’ is not attested as a quotative, though, so perhaps we should propose a direct path from simile to hesitative. The opposite development is attested in Udi and Agul (Ganenkov et al. 2010: 111).

The frequency count has established *a* ‘uh’ as another important hesitative. Previously, it was transcribed only sporadically in the Kalamang corpus due to a lack of awareness. Like *huh?*, a questioning interjection for other-initiation of

repair, a ‘uh’ might be a universal hesitative with cross-linguistically very similar pronunciation, namely with the articulators in near-neutral position (Dingemanse et al. 2014). Similar forms have been discussed for English, Dutch and German (Clark & Tree 2002, De Leeuw 2007). If *uh* is truly universal, then this form, like *huh?*, might be the result of convergent cultural evolution, reflecting “selective pressure towards the evolution of common optimised forms” (Dingemanse et al. 2014: e78273).

The differences between speakers in frequency of hesitative use can likely again be explained by the topic of conversation. Like for the placeholders, Hair has a very high frequency, likely again due to the fact that he is constantly scraping his memory for medicines to discuss.

The breakdown of the two hesitatives into pronunciation variants and by speaker can inform choices in the development of a standardised spelling by choosing the spelling that is closest to the most common pronunciation. While [nein] is the most common pronunciation for *nain* ‘like’, it is only used by one speaker in the subset of the corpus. If we collapse lengthened and non-lengthened pronunciations, there are 19 instances of [nein], all by the same speaker, while there are 11 instances of [nain] by all four speakers that use this hesitative. It seems therefore best to spell this hesitative as *nain*. A ‘uh’ mainly has the pronunciation variants [a], [e] and [ə] and the consonants [m] and [h]. The latter only occurs in interjections in Kalamang, not in other word classes. The pronunciation [a] is used by four of the five speakers that use this hesitative, so *a* seems to be the best spelling for ‘uh’ based on this sample. While native speakers might not be interested in the spelling of a hesitative like *a* ‘uh’, since they would not write it anyway, they certainly write the conventionalized form *nain* ‘like’. In addition, this kind of analysis helps choosing the best spelling when transcribing recordings for linguistic use, a task for which it is common to train native speakers (Chelliah 2013: 62). However, to make a final decision about the spelling of these items, one would need to look at more data first, as this analysis is based on very few items and four or five speakers.

4 Conclusion

Despite the small sample of the Kalamang corpus used for this study, it has contributed to further understanding of Kalamang fillers. We have seen that the ratio of placeholders is higher in this sample than in the entire Kalamang corpus, which is partly due to better transcription and glossing, but may also have been affected by the choice of genre (conversations, excluding narratives) and by individual speakers. This stresses the need for (language documentation) corpora

that contain many genres and speakers (Woodbury 2003: 46–47), and for being explicit about these parameters in studies that use (parts of) these corpora.

The higher ratio of placeholders in the current sample may also have been affected by the topics represented. The text with most hesitatives and placeholders is about a topic that one of the speakers did not find so easy to talk about, namely medicines made from plant roots (conv20). Since this was also one of the longer texts, this might have affected the ratio of fillers, but note that the ratio of placeholders in the sample is still higher than what has been found in other languages when removing this one speaker, namely 8.99 per 1000 words. It can be hypothesized that talking about less familiar topics, or a topic one is not well-prepared for (in the case of the root medicine conversation) triggers the use of more fillers. This could be further investigated experimentally, for example with a picture matching task with pictures of familiar and unfamiliar objects. Earlier research (Rose 2021) found no use of placeholders at all in rehearsed genres and very few placeholders in stimulus-based fiction, also indicating that familiarity with the topic or planning plays a role.

The refined coding of hesitatives in the current sample has revealed that a ‘uh’, analysed before as a “filler interjection” (Visser 2022: 148), is a quite common hesitative. This is not surprising, as such a hesitative may be a universal feature of languages (Dingemanse et al. 2014).

Another finding is that some speakers seem to use placeholders more often than hesitatives. Nurmia, Hawa and Fajaria all use more placeholders than hesitatives, while it is the other way around for Hair, Sabtu and Kamarudin (see Table 4). At least for Russian, there is data suggesting that hesitatives, especially non-lexical ones, are much more frequent than placeholders (Kibrik & Podlesskaya 2009).¹⁴ There is no good explanation for this difference at the moment, but note that all women use more placeholders, while all men use more hesitatives. So for now, we can only hypothesise that there is a gender difference in the relative frequency of placeholders and hesitatives.

Finally, analysing the pronunciation of hesitatives and other “peripheral” words like interjections, which often show more variation in pronunciation than other words, can inform the development of a standardised spelling in language documentation projects.

¹⁴Pakendorf (2025 [this volume]) also finds inter-speaker variation in the frequency of placeholders versus hesitatives in her Negidal corpus, but all speakers use placeholders more frequently than hesitatives. Note, however, that these results are not entirely comparable to the Kalamang data because she only regards the placeholder versus hesitant uses of a single filler, and disregards the ‘uh’ hesitative.

Table 4: Placeholder and hesitative ratio per 1000 words per speaker.

Speaker	Placeholder ratio	Hesitative ratio	Words
Hair	28.42	92.36	563
Nurmia	12.80	5.43	2579
Hawa	10.74	1.07	931
Fajaria	10.06	7.32	1093
Sabtu	8.33	13.89	360
Kamarudin	3.95	14.82	1012
Samsia	0	0	246
Bini	0	0	563

Acknowledgments

I would like to thank the editors of this volume for several rounds of very careful reviewing, which greatly improved the quality of this chapter.

Abbreviations

ANA	anaphoric demonstrative	<i>lat</i>	ablative/allative
AT	attributive	LOC	locative
ATTEN	attenuative	NEG	negator
CLF_LEAF	classifier for thin flat entities	NFIN	nonfinal
		OBJ	object
COM	comitative	OBJQNT	object quantifier
DIST	distal	PL	plural
EMPH	emphasis	PLNK	predicate linker
EXCL	exclusive	POSS	possessive
EXIST	existential	PROX	proximal
FOC	focus	SG	singular
IAM	iamitive	SIM	similative
INCL	inclusive	TAG	tag
INT	interjection	TOP	topic
IRR	irrealis		

References

- Billings, Blaine & Bradley McDonnell. 2025. Fillers in the gray: Indeterminacy in their use in everyday conversation. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 211–250. Berlin: Language Science Press. DOI: ??.
- Chelliah, Shobhana. 2013. Fieldwork for language description. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 51–73. Cambridge: Cambridge University Press.
- Cheung, Lawrence Yam-Leung. 2015. Uttering the unutterable with *wh*-placeholders. *Journal of East Asian Linguistics* 24. 271–308.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- De Leeuw, Esther. 2007. Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19(2). 85–114. DOI: 10.1017/S1470542707000049.
- Dingemanse, Mark, Francisco Torreira & N. J. Enfield. 2014. Is “huh?” a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS One* 8(11). e78273. DOI: 10.1371/journal.pone.0078273.
- Döhler, Christian. 2025. *That* placeholder in Komnzo. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 275–313. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697589.
- Egorova, Maria A. 2018. Discourse marker *tipa* according to the data of Russian National Corpus: Its origin, semantics and pragmatics. In *Computational linguistics and intellectual technologies. Papers from the annual international conference “Dialogue”*, 177–191.
- Enfield, N.J. 2003. The definition of *what-d’you-call-it*: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.

- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 33–66. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.03hay.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University press.
- Kibrik, Andrei Aleksandrovich & Vera I. Podlesskaya (eds.). 2009. *Rasskazy o snovidenijax: Korpusnoe issledovanie usnogo russkogo diskursa [night dream stories: A corpus study of spoken Russian discourse]*. Moscow: Jazyki Slavjanskix Kul'tur.
- Kluge, Angela. 2017. *A grammar of Papuan Malay* (Studies in Diversity Linguistics 11). Berlin: Language Science Press. DOI: 10.5281/zenodo.376415.
- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hözl & Thomas Payne (eds.), *Tungusic languages: Past and present*, 199–225. Berlin: Language Science Press.
- Klyachko, Elena. 2025. Placeholders versus general extenders in Tungusic languages. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 41–74. Berlin: Language Science Press. DOI: ??.
- Laserna, Charlyn M., Yi-Tai Seih & James W. Pennebaker. 2014. *Um... Who like says you know*: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology* 33(3). 328–338.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- Pakendorf, Brigitte. 2025. A corpus-based investigation of fillers in Negidal (Northern Tungusic). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 75–129. Berlin: Language Science Press. DOI: 10.5281/zenodo.1569759.
- Palacios Martínez, Ignacio M. & Paloma Núñez Pertejo. 2015. “Go up to miss thingy”. “He’s probably like a whatsit or something”. Placeholders in focus. The differences in use between teenagers and adults in spoken English. *Pragmatics* 25. 425–451. DOI: 10.1075/prag.25.3.05pal.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.

- Podlesskaya, Vera I. & Andrei Aleksandrovich Kibrik. 2006. Korrektsiya v ustnoi russkoi monologicheskoi rechi po dannym korpusnogo issledovaniya [Correction in Russian spoken monologues: A corpus study]. *Russkii yazyk v nauchnom osveshchenii [Russian Language and Linguistic Theory]* 12. 7–55.
- Ponsonnet, Maïa. 2025. Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 315–340. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697591.
- Rose, Françoise. 2021. Mojeño Trinitario fillers and placeholder. Presentation at Atelier morphosyntaxe, DDL, Lyon.
- Seraku, Tohru. 2024. Placeholders in crosslinguistic perspective: Abilities, preferences, and usage motives. *Linguistics* 62(2). 1–26. DOI: 10.1515/ling-2023-0068.
- Usher, Timothy & Antoinette Schapper. 2022. The Greater West Bomberai language family. *Oceanic Linguistics* 61(1). er1–er59. DOI: 10.1353/ol.2022.0004.
- Ventayol-Boada, Albert. 2025. Copula or placeholder: A qualitative and quantitative analysis of *λe-* in Kolyma Yukaghir discourse. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 131–168. Berlin: Language Science Press. DOI: ??.
- Visser, Eline. 2020. *The Kalamang collection: An archive of linguistic and cultural material from Karas*. Lund: Lund University Humanities Lab Archive. <http://hdl.handle.net/10050/00-0000-0000-0003-C3E8-1>.
- Visser, Eline. 2022. *A grammar of Kalamang* (Comprehensive Grammar Library 4). Berlin: Language Science Press. DOI: 10.5281/zenodo.6499927.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. *Language Documentation and Description* 1. 35–51.
- Zhao, Yuan & Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. In *Proceedings of DiSS'05: Disfluency in spontaneous speech workshop*, 179–182. https://www.isca-speech.org/archive_open/diss_05/dis5_179.html.

Chapter 8

That placeholder in Komnzo

Christian Döhler

Berlin-Brandenburg Academy of Sciences and Humanities

Speakers of Komnzo have a number of conventionalised devices for situations of disfluency. In addition to silent pauses, there are hesitant and placeholder fillers. This contribution places a focus on the placeholder *bäne* which is a pronoun, or pro-form in the language. The chapter consists of a description of the form, distribution, functions, functional extensions, frequency, and multimodality of *bäne*. It therefore contributes to the emerging typology of placeholders.

keywords: Yam languages, Southern New Guinea, fillers, placeholders, demonstratives

1 Introduction

This chapter describes and analyses fillers in Komnzo, a language of the Yam family spoken in the south-west of Papua New Guinea.¹ Fillers are linguistic devices that are used in situations of disfluency, i.e., they are filling a silent pause. They can be divided into hesitant fillers (e.g. English *uhm*, Komnzo *a*) and placeholder fillers (e.g. English *whatchamacallit*, Komnzo *bäne*). The former are non-referential and not syntactically integrated, while placeholders are referential and syntactically integrated (Hayashi & Yoon 2010).²

While I touch on hesitant fillers only in passing, the main focus of the chapter is on the placeholder *bäne/baf*. I show that this placeholder is best analyzed as a pro-form which has developed from a medial demonstrative. I argue that the use of *bäne/baf* goes well beyond filling a silent pause. It is used intentionally with

¹ISO 639-3: tci, Glottocode: komn1238

²Previous classifications have defined hesitant fillers as non-lexical but nonetheless conventionalized sounds, while placeholder fillers are lexical items (cf. Amiridze et al. 2010).



communicative goals such as signalling a taboo context, or discourse managing goals such as gaining the floor. Moreover, I show that certain inflections of *bäne* have evolved into conventionalized connectors for adverbial clauses.

In the remainder of this section, I introduce the sociolinguistic situation (§1.1), the text corpus (§1.2), and some typological features of the languages (§1.3 and §1.4). §2 provides an overview of hesitant and placeholder fillers in Komnzo. The main body focusses on the placeholder *bäne/baf*, for which I describe its form (§3.1), distribution (§3.2), functions (§3.3), functional extensions (§3.4), and multimodal aspects (§3.5). After addressing the problems in measuring the frequency of *bäne/baf* (§3.6), I close with some final comments (§4).

1.1 Sociolinguistic background

Komnzo is a small language even by the standards of Papua New Guinea, where language communities tend to be rather small. Komnzo is spoken by approximately 200–250 speakers in the villages of Rouku and Morehead Station. Genetically, the language belongs to the Tonda subgroup of the Yam languages. Figure 1 shows a map of the language family.

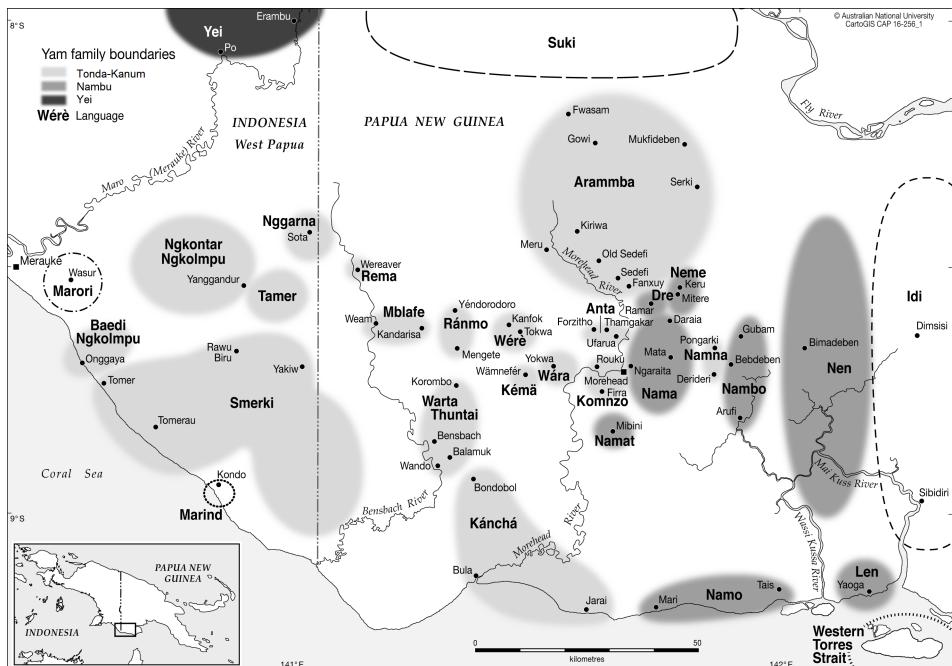


Figure 1: The Yam language family

Komnzo speakers live in a highly multilingual language ecology. Due to a marriage pattern of sister-exchange with exogamous groups based on clan, place, and – epiphenomenally – on language variety, virtually all children grow up with at least two languages. In reality, the portfolio of most children includes 4-5 languages by the time they reach adulthood.

Komnzo speakers live in a small-scale traditional society, i.e. what has been called “society of intimates” (Givón 2018, Givón & Young 2002). In this type of social setting all generic knowledge is shared and almost all daily interactions take place between individuals who have known each other for a long time. This results in a large degree of common ground, thus, leading to higher informational predictability in face-to-face conversation.

1.2 Text corpus and methodology

The data discussed in this chapter is based on recordings made between 2010 and 2015, archived as (Döhler 2021). The corpus used here comprises around 12 hours of various text genres, including both natural and stimuli-based narratives and conversations (Table 1). The overall size is around 55,000 word tokens, which makes the Komnzo text corpus a typical “language documentation corpus” (Mosel 2012).

Table 1: Corpus overview

Text type	hh:mm:ss
Conversations	01:01:55
Conversational tasks	01:49:51
Narratives	06:40:18
Procedural texts	02:11:36
Public speeches	00:42:38
Total	12:26:18

All examples are referenced with a source code of the following format: [tci-YYYYMMDD-NN SSS ##]. The first part identifies the transcription file. Each session and the included files start with the ISO 639-3 code for Komnzo: tci. Next comes the date of the recording (YYYYMMDD) and the session number on that date (NN). The second part identifies the example within the transcription file. Transcription tiers are sorted by speaker (SSS). Intonation units on the respective

transcription tiers are numbered (##). Thus, example (1) in this chapter has the source code [tci20150906-10 ABB 303–306]: it is the tenth recording session on September 6th, 2015, the speaker is Abia Bai (ABB), and on the speaker's text tier, the example shows the annotation units 303–306. The corpus of transcribed and interlinearised texts has been archived as (Döhler 2021).³

Most examples in this chapter include a figure showing the wave file, the pitch contour, and the transcription. Pauses in the example were measured in ELAN and are shown with three dots on the text tier (...) and in milliseconds on the gloss tier (ms). For producing the figures, the wav-files were exported from ELAN, normalized and converted to mono track files in Audacity, and finally imported to Praat.⁴ All processed wav-files, Praat pictures, and video screenshots can be downloaded from: <https://zenodo.org/doi/10.5281/zenodo.12032997>.

1.3 Typological overview

Komnzo is a double-marking language, in which the verb indexes core arguments and noun phrases are flagged for case. The case marking is organised in an ergative-absolutive system. In addition to four core cases (absolutive, ergative, dative, possessive), there are 13 semantic cases. Verbs are by far the most complex part of speech in the language. Verbs mark person, number and gender of up to two participants, 18 TAM categories, valency, directionality and deictic status. Complexity lies not only in the amount of grammatical categories that can be expressed morphologically, but also in the way these categories are encoded (Döhler 2018: 175ff.). This is best described by the term “distributed exponence” (Caballero & Harris 2012, Carroll 2016), a subtype of multiple exponence.

This aspect of the language is not the topic of this chapter, but it has a practical effect for the presentation of example sentences, in that I do not show the morpheme segmentation of verbs. Instead, I apply the word-and-paradigm approach (Matthews 1974): In the morpheme tier, I separate the verb stem from affixal material by placing it between \slashes/. In the gloss tier, I list the relevant grammatical categories (argument structure, TAM, directionality) followed by the lexeme translation.

³While (Döhler 2021) contains a zipped file of all transcriptions, the audio-visual footage of each session can be found under: <https://zenodo.org/communities/komnzo>

⁴ELAN (version 6.7): <https://archive.mpi.nl/tla/elan>, Audacity (version 2.4.2): <http://audacity.sourceforge.net/>, Praat (version 6.4.04): <https://www.fon.hum.uva.nl/praat>

1.4 Demonstratives

Before we proceed it is worth defining the notion of demonstratives adopted in this chapter and presenting some basics about the system of demonstratives in Komnzo. As a point of departure, I follow Diessel (1999: 2ff.) in assuming that the most basic function of demonstratives is a spatial (or situational) use, but see Himmelmann (1996) and De Mulder (1996) for a critique of this view. Based on this functional definition, we can identify the forms given in Table 2 as demonstratives.

Table 2: Demonstratives in situational uses

PRONOUN /		ADVERB		CLITIC
	DETERMINER	NEUTRAL	ALLATIVE	ABLATIVE
PROX	<i>zane</i>	<i>zä</i>	<i>zbo</i>	<i>zba</i>
MED	(<i>bäne</i>)	<i>bä</i>	<i>bobo</i>	<i>boba</i>
DIST	(<i>ane</i>)	<i>fä</i>	<i>fobo</i>	<i>foba</i>
Q	<i>mane</i>	<i>mä</i>	<i>mobo</i>	<i>moba</i>

The demonstrative paradigm is organized in a four-way split into proximal, medial, distal, and (spatial) interrogative.⁵ Table 2 shows that the distinction is signalled by the initial consonant: /z/ for proximal, /b/ for medial, /f/ for distal, and /m/ for interrogative. In this aspect, demonstratives are related to person deixis, i.e. personal pronouns. The first person singular pronouns start with /nz/ (*nzä* 1SG.ABS)⁶, second person pronouns with /b/ (*bä* 2SG.ABS), and third person pronouns with /f/ (*fi* 3.ABS). The implications of this link, especially between second person and the placeholder *bäne*, are addressed in §3.1.

Syntactically demonstratives belong to different parts of speech, which align with Diessel's classification: pronouns, determiners, adverbs, and identifiers (Diessel 1999). The elements shown in the leftmost column can be used both adnominal and pronominal, i.e. they function as determiners (*zane mni-me* [PROX fire=INS] 'with this fire') and pronouns (*zane=me* [PROX=INS] 'with this one'). The elements in the middle of the table function as adverbs (*zä* 'here',

⁵The latter category is glossed as Q, and it is used for questions that pertain to space: *mane* 'which one', *mä* 'where', *mobo* 'whither', and *moba* 'whence'.

⁶Note that the Komnzo first person pronoun starts with /nz/, and not /z/. This is different in closely related varieties such as Wära, Anta, and Wèré where first person pronoun all begin with /z/.

zbo ‘hither’, *zba* ‘hence’). The elements in the right column are verbal proclitics, and their most frequent use is as part of a presentational (or identificational) construction.⁷

In the remainder of this section, I will focus on the forms in the leftmost column of Table 2. There are two elements in the table, namely *bäne* and *ane*, which formally belong in this paradigm, but are never used situationally, i.e. they do not point to something in space. The first is the anaphoric *ane*, for which there is evidence that it has developed from an older form *fane* (Döhler 2018: 110ff.). *Ane* no longer has the (distal) spatial reference that is suggested by its position in the paradigm. Instead, it is used anaphorically for referents or sometimes for a whole proposition that has been established in the preceding discourse, i.e. it is used for “tracking” (cf. Himmelmann 1996). This is shown with both tokens in (1). Note that, unlike *ane*, the proximal *zane* can be used both anaphorically and cataphorically. Syntactically, *ane* functions as a pronoun (1) and adnominally as a determiner (3).

- (1) *ruga nzmär=me, yti ane thf\konzr/mth fof, yti, ane=me*
pig grease=INS PN DEM 3PL>3PL:PST:DUR\speak EMPH PN DEM=INS
za\nänzütham/ath
3PL>3SG.F:PST:PFV\paint
'with pig grease. They were really calling **this** *yti*. They painted her **with** this.' [tci20150906-10 ABB 303-306]

The second element is the placeholder *bäne/baf*. It has lost the spatial function that is suggested by its position in the paradigm (medial). Moreover, *bäne/baf* is almost never used adnominally, i.e. as a determiner. Such examples were not only assessed as ungrammatical during elicitation, but the corpus also points in this direction. Example (16) is the only example (out of more than 700 tokens in the corpus) that could be argued to be an adnominal use. This confirms Hayashi & Yoon (2010), who mention that the majority of placeholder uses of demonstratives in their comparative study of Japanese, Korean and Mandarin were pronominal rather than adnominal. Note also that *bäne/baf* is never used anaphorically in Komnzo. For further elaboration on *bäne/baf*, I refer the reader to §2.2 and §3.

⁷The proclitics can attach to any inflected verb. In the presentational construction, they attach to the copula which follows the main verb of the clause (cf. Döhler 2018: 109ff., 288).

2 Overview of hesitators and placeholders

This section describes hesitative fillers (§2.1), placeholder fillers (§2.2) and other fillers (§2.3). In the section on placeholder fillers, I introduce three devices: the placeholder *bäne/baf*, the light verb *-rä* ‘do’, and the manner demonstrative *nima*. These can be used to replace nominal elements, verbs or entire sections of discourse.

2.1 Hesitative fillers

Disfluency in the speech of Komnzo speakers can manifest itself as a stretch of silence, or it can be filled by a hesitator.⁸ Hesitators are usually pronounced as open vowels of variable length, often [æ]~[e]. The hesitator is usually followed by a very short pause before fluent speech continues. Examples from the Komnzo text corpus confirm what has been written about disfluencies elsewhere; e.g., Clark & Fox Tree (2002) schematize disfluencies into three phases: first, a *suspension* of fluent speaking; secondly, a *hiatus* in speaking, which may contain a stretch of silence or a hesitative filler (accompanied by other collateral actions like gestures); and thirdly, a *resumption* of fluent speaking.

Example (2) shows a clause initial hesitator, produced as [æ]. The pause between the hesitator and the following NP *no mni* is very short (70ms), the same length as the pause between the clause and the postposed NP *kafsin* (cf. Figure 2).

- (2) ä ... no mni f=|rä/ ... kafsi=n
 HES (70ms) water hot DIST=3SG.F:NPST:IPFV\be (70ms) cup(E)=LOC
 ‘uh there was tea there in a cup.’ [tci20120924-01 TRK 44]

Another hesitator is shown in example (3), produced as [e], which is followed by a pause and a relative clause. The pause following the hesitator is much longer (320ms) than the pause in the previous example, but relatively speaking, it is still very short (cf. Figure 3). This becomes evident when we compare it to the pause following the placeholder *bafen* in the same example, which is twice as long (660ms).

- (3) ane fam fof ḷa|rä/r a ... monme
 DEM thought EMPH 3SG:NPST:IPFV\do HES (320ms) how
 san\thb/ath bobo baf=en ... sel=en.
 3PL>3SG.M:PST:PFV:VENIT\put_in MED:ALL PH=LOC (660ms) cell(E)=LOC
 ‘He is thinking of, uhm, how they put him into the whatchamacallit, into

⁸I use the term “hesitator” and “hesitative filler” interchangeably.

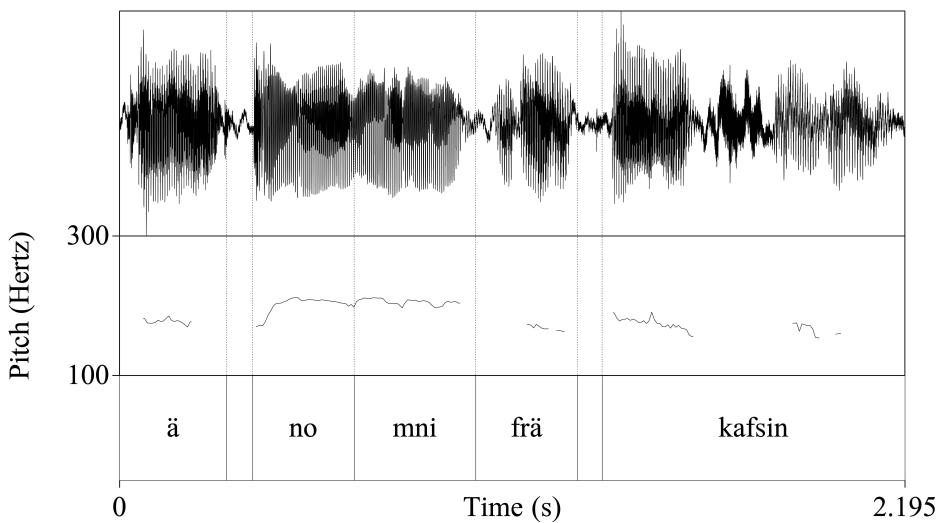


Figure 2: Audio analysis of [tci20120924-01 TRK 44]

the cell.'

[tci20111004 RMA 414–416]

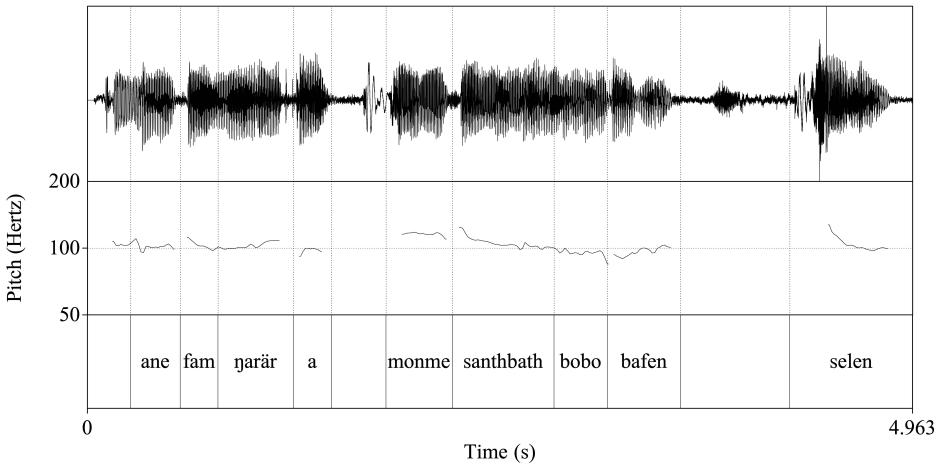


Figure 3: Audio analysis of [tci20111004 RMA 414–416]

2.2 Placeholder fillers

There are three kinds of placeholder fillers in Komnzo: the placeholder *bäne/baf*, the light verb *-rä* ‘do’, and the manner demonstrative *nima*. These can be used to replace nominal elements, verbs or entire sections of discourse. I follow Hayashi & Yoon (2006) in their definition of placeholders as referential expressions that are used as a substitute for a specific lexical item. Under this definition, placeholders occupy “a syntactic slot that would have been occupied by the target word” (2006: 490). For the present chapter, I expand the definition to include larger units, e.g. a whole clause or a proposition.

2.2.1 The placeholder *bäne/baf*

The word that follows the above definition most closely is the placeholder *bäne/baf*. We have seen a rather typical example of the placeholder in (3), in which the speaker has problems finding the appropriate word, and finally uses an English loan *sel* ‘prison cell’. The placeholder is flagged with the locative case (=en), and after a short pause the speaker produces the target *sel* with the appropriate case flag. Thus, the placeholder *bäne* is used pronominally. We will see in §3 that there are a number of functional extensions of this pattern.

Before moving on to introduce the other types of placeholders, I want to point to a recent change in my analysis of *bäne*. Formally, *bäne* patterns with the demonstratives (cf. Table 2), and based on its initial consonant /b/ it belongs in the medial category. Such a link is not surprising from a cross-linguistic perspective. Hayashi & Yoon (2006) report for Japanese, Korean, Mandarin and Indonesian that certain demonstratives also function as placeholders. This is akin to the analysis that I have adopted in the past, i.e., there is a distinction between the demonstrative *bäne* and the placeholder *bäne* (Döhler 2018: 112). The analytic criteria for setting up this distinction were based on prosody and syntax. For example, a break in the intonation contour through a short pause signals a disfluency situation and following from this such examples were analysed as placeholders (e.g. *bäne* (.) *kabe* ‘who’s-that ... the man’). If there is no disfluency situation, *bäne* is analysed as a demonstrative (e.g. used adnominal: *bäne kabe* ‘that man’). Based on a rigorous inspection of corpus examples, I have now abandoned this analysis.

Despite its etymological origin in the system of demonstratives, I have come to the conclusion that *bäne* is not used as an (exophoric/situational) demonstrative at all. The prosody of all inspected tokens points to disfluency situations, and what seemed like adnominal uses of a demonstrative can be accounted for by

assuming a nominal compound in which the first noun is filled by a placeholder, as in example (4) below. I put this here as a first introduction to *bäne*, and as a disclaimer, before we delve into the details in §3.

- (4) *kwas nima kam zä\kwthef/a bäne zawe ... töna*
brolga like_this back SG:PST:PFV\turn PH side (1350ms) high_ground
zawe
side

'The brolga turned its back to the what's-that side ... to the land side.'

[tci20130923-01 ALB 51-52]

2.2.2 The manner demonstrative *nima*

The second element in Komnzo to be discussed here is the manner demonstrative *nima*, which I translate as 'like this'. Sometimes it occurs with the instrumental case as *nima=me* [like_this=INS].

This word is not a placeholder per se, but it can be used as a placeholder in certain contexts. *Nima* is used to further elaborate on the manner in which some event was carried out. The manner component is understood from context in most corpus examples, i.e. it is often not spelled out. Additionally, it can be accompanied by a gestural component, as in (4) above, in which the speaker turns her body away to reenact the movement of the brolga. However, the manner component can also be verbalized, which provides a bridging context for the placeholder use of *nima*. This is also a bridging context for another use of *nima*, namely as a quotative marker. Especially in this latter function, *nima* shares some characteristics with English *like*. In its placeholder use, but also in the use as a quotative marker, *nima* is always followed by a pause, as in (5).

There are two differences to the placeholder *bäne*. First, *nima* never substitutes a nominal element. As a manner demonstrative, it is used in place of a more complex event, often a stretch of discourse. What follows *nima* in speech is usually a more fine-grained elaboration. In example (5), the speaker elaborates on the people who were present in the situation. In this placeholder use of *nima*, there is a long pause (1060ms) followed by a whole clause (cf. Figure 4).

- (5) *wati, ä\kwa/thake nima ... sitau=aneme afa*
then 1PL>3PL:PST:IPFV\cut_meat like_this (1060ms) PN=NSG.POSS father
kwark b=ya\r/a nafanm ä\kwa/ne,
deceased MED=3SG.M:PST:IPFV\be 3NSG.DAT 1PL>3DU:NPST:IPFV\cut_meat

nzenm=wä

1NSG.DAT=EMPH

‘Then, we cut the meat **like this** ... Sitaú’s late father (and his wife) were there. We cut some for them, and also some for us.’

[tci20120821-02 LNA 95–97]

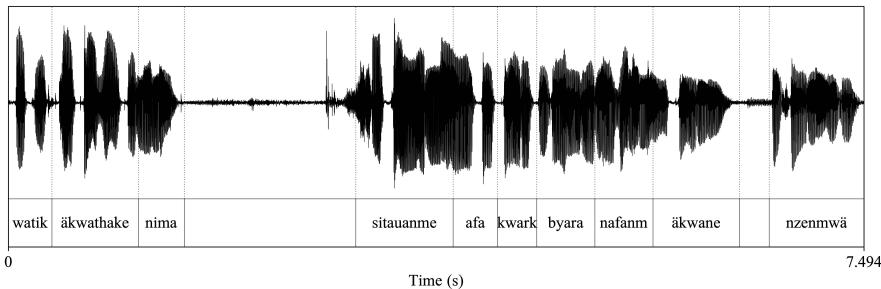


Figure 4: Audio analysis of [tci20120821-02 LNA 95–97]

Secondly, *nima* can be used anaphorically and cataphorically. Example (6) shows a cataphoric use of *nima*. In the text example, a man suddenly realises that his visiting relative has actually come as the vanguard of a group of headhunters. The cataphoric use (6) and the placeholder use (5) of *nima* can be distinguished by the fact that only the latter occurs in disfluency situations. In the placeholder example (cf. Figure 4), there is a significantly longer pause than in the cataphoric example (cf. Figure 5): 1060ms versus 280ms.

- (6) *fi miyatha sf|rä/rm nima ... zan=r zä*
 3.ABS knowledgeable 3SG.M:PST:DUR\be like_this (280ms) kill=PURP PROX
zf swan\yak/
 EPS 3SG.M:PST:IPFV\come
 ‘He knew it: he had come here to kill (people).’ [tci20111119-01 ABB 98–98]

(7) is a typical example of anaphora with *nima*. The speaker makes a resumptive comment summarizing a description of how to catch fish in the swamp. There is no pause or any other sign of disfluency in the audio of this example.

- (7) *zra=ma trikasi nima=me fof \rä/*
 swamp=CHAR story like_this=INS EMPH 3SG.F:NPST:IPFV\be
 ‘The story about the swamp is just like this.’ [tci20120922-09 DAK 48]

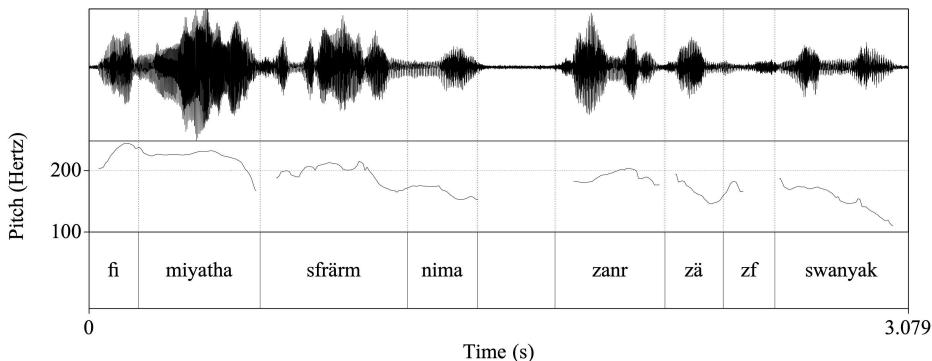


Figure 5: Audio analysis of [tci20111119-01 ABB 98–98]

2.2.3 The light verb *-rä*

The third element that can be used as a placeholder is the light verb *-rä* ‘do’.⁹ Komnzo has a number of light verbs, e.g. *fiyoksi* ‘make’, *-kor* ‘become’, *wäsi* ‘happen’.¹⁰ These are used in light verb constructions in which the verb carries the inflectional information, while a nominal element expresses the verbal semantics (cf. Döhler 2018: 304). A typical context for such constructions is the integration of borrowed nouns (from a verb in the source language), as in (8) below.

- (8) *no=r bobo zek krä|r/é*
 water=PURP MED:ALL check(E) 1SG:IRR:PFV\do
 ‘I would (go) there and check for water.’ [tci20130903-03 MKW 146]

There are certain contexts in which a speaker uses the verb *-rä* not in a light verb construction, but because s/he has trouble finding the correct lexical entry. In corpus examples of this type, there is usually a short pause after the light verb, and then the correct full verb follows, sometimes the entire clause is repeated with the full verb. We can see this in examples (9) and (10) below. In both examples the light verb and the full verb carry the identical inflectional pattern in terms of alignment and TAM categories.

In example (9), the speaker describes the final stage of the ritual destruction of a grave site. Since he has trouble finding the correct verb for ‘levelling the

⁹The verb ‘do’ is heterosemous (cf. Lichtenberk 1991 and Evans 2010) with the copula verb ‘be’. In one type of inflectional pattern *-rä* means ‘do’, and in another pattern it means ‘be’ (cf. Döhler 2023).

¹⁰Some verbs lack an infinitival form. These are given here with a hyphen.

ground', he pauses (550ms), and then uses the light verb. After a longer pause (1800ms), he continues with the full verb *frmzsi* 'prepare, straighten' (cf. Figure 6).

- (9) *wati ane bad kwot we ... zfrä/rme we*
 then DEM ground(ABS) properly also (550ms) 1PL>3SG.F:PST:DUR\do also
 ... *zwa\frmnzr/me* *nima*
 (1800ms) 1PL>3SG.F:PST:DUR\straighten like_this
 'Then we were properly doing the ground ... we were levelling it like this.'
 [tci20120805-01 ABB 831–832]

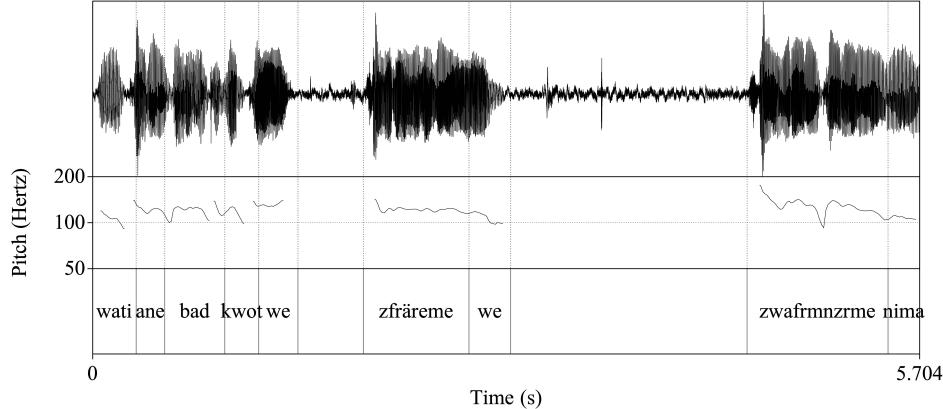


Figure 6: Audio analysis of [tci20120805-01 ABB 831–833]

Example (10) comes from the afterword of a recording. The speaker explains that his story is not a fictional story, but a real story. He substitutes the target word *kwthenzzi* 'change, turn' with the light verb *-rä*. In this example, there is no pause preceding the light verb, and the following pause (550ms) is rather short (cf. Figure 7).

- (10) *kabe zokwasi aha nzürna trikasi za|r/ath* ...
 man story yes spirit story 3PL>3SG.F:PST:PFV\do (550ms)
za|kwthef/ath
 3PL>3SG.F:PST:PFV\change
 'A real story. Yes, they made it into a spirit story ... they changed it.'
 [tci20111119-06 MAB 146–147]

While Komnzo utilizes light verbs to function as placeholders, there are other Papuan languages that have dedicated placeholder verbs, for example *məgi-* 'to do whatever' in Manambu (Aikhenvald 2008: 576).

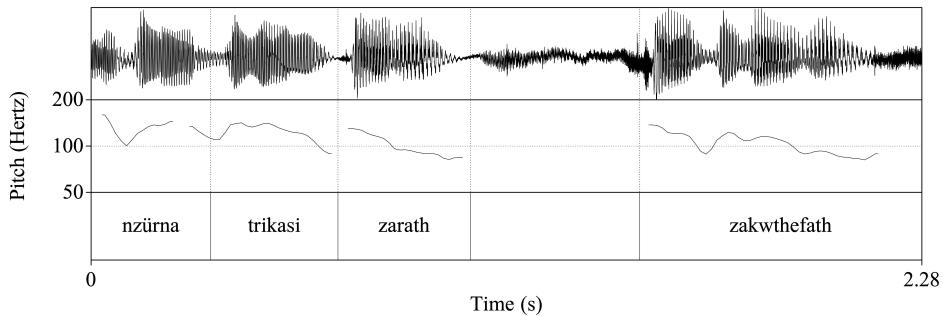


Figure 7: Audio analysis of an excerpt of [tci20111119-06 MAB 146–147]

2.3 Other fillers

There is another element in Komnzo that can be used as a filler, but its main function is connected to narrative style rather than disfluency. Especially the passing of time or the distance that one has travelled is often indicated with a long stretched *e*, realized as [ɛ:], which I gloss as ‘until’. While this function is not a case of disfluency, the long stretched *e* is sometimes used like a hesitator, and it is the specific context that facilitates this use. In most corpus examples, the long stretched *e* is followed by a place name, as in (11) and (12). But there is nothing in the prosody that would differentiate its use for narrative style from a disfluency situation. The fact that the speaker was indeed searching for the place name in the two examples below, only became clear during the transcription of the texts, in which the speakers themselves explained it in this context.

The stretched *e* can be quite short. In example (11) it is 200ms long (cf. Figure 8), not much longer than the pause between the clause and the postposed noun (180ms).

- (11) *fi we krän\brim/ e masu ... garda=me=nzo*
 3.ABS also SG:IRR:PFV:VENIT\return until PLN (180ms) canoe=INS=ONLY
 ‘He returned to Masu by canoe.’ [tci20100905 ABB 81–82]

It can also be very long, as in example (12), where it is 1600ms and takes up almost half of the length of the intonation unit (cf. Figure 9).

- (12) *foba fof kre\far/é e wi\yak/ e büdisn*
 DIST:ABL EMPH 1SG:IRR:PFV\set_off 1SG:NPST:IPFV\walk until PLN
 ‘From there I set off and walked until Büdisn.’ [tci20130903-03 MKW 23]

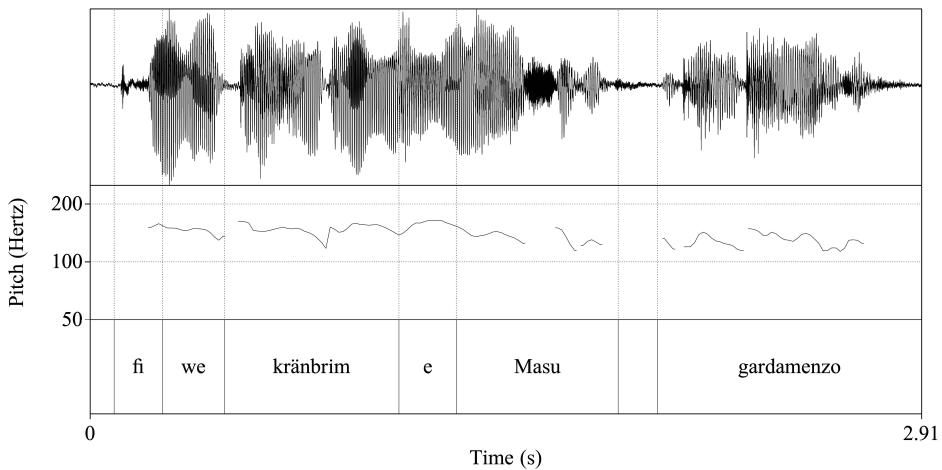


Figure 8: Audio analysis of [tci20100905 ABB 81–82]

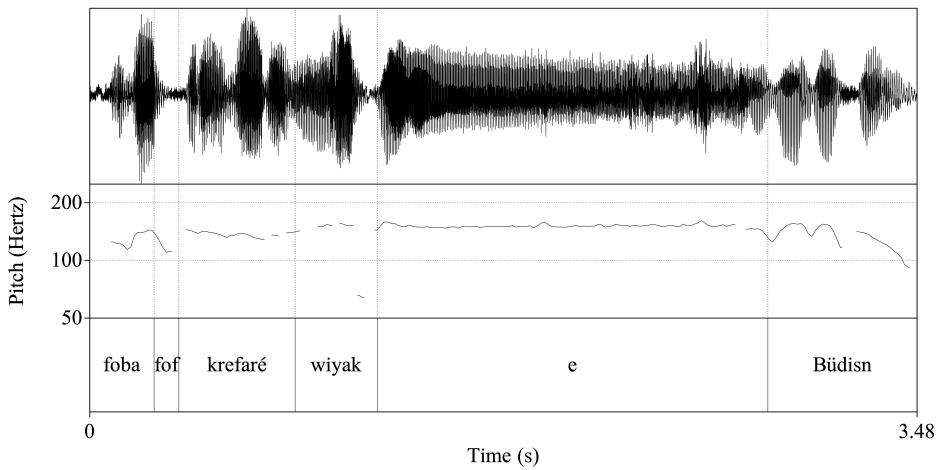


Figure 9: Audio analysis of [tci20130903-03 MKW 23]

3 The placeholder *bäne/baf*

This section describes the placeholder *bäne/baf* in its form (§3.1), syntactic distribution (§3.2), its various functions (§3.3) and extensions thereof (§3.4). I show three examples of accompanying gestures in §3.5 and discuss the problems of measuring the frequency of the placeholder in §3.6.

I adopt the terminology of “delayed constituent” and “target” for the word or larger unit which the placeholder substitutes. In examples, I print the placeholder (and sometimes a larger unit to which the placeholder belongs) in **bold font** and the delayed constituent or target in underlined font, as in (13) below.

- (13) *zöbthé zwa\wärez/é bäre=me ... kofä tot=me*
first 1SG>3SG.F:RPST:PFV\aim PH=INS (280ms) fish spear=INS
‘First I aimed at it **with the whatchamacallit** ... with the fish spear.’
[tci20130905-02 MKW 41–42]

3.1 Form

As I have argued in §2.2.1, the word *bäre* belongs formally in the paradigm of demonstratives, more specifically in the medial category (cf. Table 2). Demonstratives in Komnzo can be used pronominally, and in this syntactic position they can be flagged with a subset of case markers, as is shown in Table 3 for the proximal demonstrative *zane*.

The placeholder *bäre* is much more active in this respect. As can be seen in Table 3, it can also be flagged for the three spatial cases (LOC, ALL, ABL) with inanimate referents. More importantly, *bäre* has a second stem, *baf*, which can be flagged for all cases that encode animate referents with the standard number distinction. In the Table, we can see some spillover of the two stems, i.e. *bäre* for animates and *baf* for inanimates, which I take as evidence that the two stems belong to the same underlying form. For example, *bäre* appears as the absolute case form for animate referents, while all other forms are built from *baf*.¹¹ Conversely, the locative case form for inanimate referents is built from *baf*, not from the expected *bäre*.

In terms of possible case flags, *bäre/baf* is not only more active than the demonstratives, but also more active than the person pronouns, which in turn cannot be flagged for cases that encode inanimate referents. In Table 3, I show the possibilities of personal pronouns with the 2SG in the rightmost column. Thus, one can say that the placeholder *bäre/baf* is the “most prototypical pronoun” in the language because it can substitute all nouns inflected for all cases. However, the agnostic term “pro-form” is more suitable, since *bäre/baf* can also replace longer stretches of discourse.

While it is clear from the paradigm in Table 2 that *bäre* has developed directly from the demonstratives, we can only speculate on the origin of the second stem

¹¹The lack of a number distinction is found with all absolute case forms, e.g. personal pronouns and nominal enclitics.

Table 3: Different case inflections

		PLACEHOLDER		DEM	PERS. PRON
	INANIM	ANIM (SG)	ANIM (NSG)	PROX	2SG
ABS	<i>bäne</i>	<i>bäne</i>	<i>bäne</i>	<i>zane</i>	<i>bä</i>
ERG	-	<i>baf</i>	<i>baf-a</i>	-	<i>bné</i>
DAT	-	<i>baf-an</i>	<i>baf-anm</i>	-	<i>bun</i>
POSS	-	<i>baf-ane</i>	<i>baf-anme</i>	-	<i>bone</i>
CHAR	<i>bäne=ma</i>	<i>baf-ane=ma</i>	<i>baf-anme=ma</i>	<i>zane=ma</i>	<i>bone=ma</i>
LOC	<i>baf=en</i>	<i>bafa-db=en</i>	<i>baf-anme-db=en</i>	-	<i>bun-db=en</i>
ALL	<i>bäne=fo</i>	<i>bafa-db=o</i>	<i>baf-anme-db=o</i>	-	<i>bun-db=o</i>
ABL	<i>bäne=fa</i>	<i>bafa-db=a</i>	<i>baf-anme-db=a</i>	-	<i>bun-db=a</i>
IC	-	<i>baf=rr</i>	<i>baf=ä</i>	-	<i>bn=rr</i>
INS	<i>bäne=me</i>	-	-	<i>zane=me</i>	-
PURP	<i>bäne=mr</i>	-	-	<i>zane=mr</i>	-
PROP	<i>bäne=kära</i>	-	-	<i>zane=kära</i>	-
PRIV	<i>bäne=mär</i>	-	-	<i>zane=mär</i>	-

baf. One hypothesis is that *bäne* first merged with the ergative case enclitic =f [ERG.SG], and then developed the case forms shown in Table 3. Also recall that – based on the initial /b/ consonant – the medial category in the deictic system is formally (and historically) related to second person pronouns. Therefore, there is a historical link between the second person pronouns, the medial category in the deictic system, and the placeholder. Synchronously, I do not analyse deixis in Komnzo as a person-based system in the sense of Keenan & Anderson (1985).

3.2 Distribution

The placeholder can substitute any noun or noun phrase, whether it be case marked or zero marked (i.e. absolute). A typical example is given in (14), where the speaker uses *bäne* followed by a very short pause of 160ms (cf. Figure 10), and then continues with the delayed constituent *zaru yawi*, on which he further elaborates.¹² *Bäne* here substitutes the whole nominal compound.

¹² *Zaru* is the candlenut tree (*Aleurites mollucana*) and its fruit is a stone fruit. The speaker refers to the hard kernel in the example.

- (14) *fi bäre ... zaru yawi ... zaru yawi mane*
 but PH(ABS) (160ms) PN nut (2200ms) PN nut which(ABS)
y\konz/rth ... nimä=wä we fof
 3PL>3SG.M:NPST:IPFV\speak (1200ms) like_this=EMPH also EMPH
 ‘But whatchamacallit ... the zaru nut ... (that) which they call zaru nut.
 (They did) the same thing (with it).’ [tci20120818 ABB 36–38]

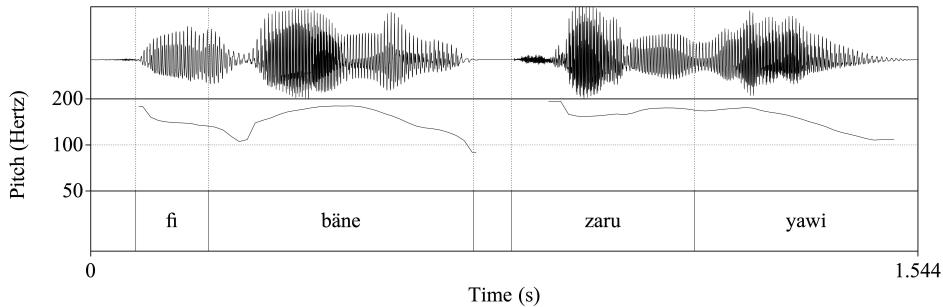


Figure 10: Audio analysis of an excerpt of [tci20120818 ABB 36]

In example (15), the placeholder occurs in the position of the head of a nominal compound ('house whatchamacallit' > '(house) platform').¹³ In the delayed constituent only the head noun is repeated, not the whole compound. There are no measureable pauses in the example, but there is a break in the pitch contour before the delayed constituent.

- (15) *mnz baf=en boba skiski=n y\rakth/kwa*
 house PH=LOC MED:ABL platform=LOC SG>3SG.M:PST:IPFV\put_on_top
 ‘I put it on top of **the house** whatchamacallit, the (house) platform.’
 [tci20111119-03 ABB 34]

In example (16), the speaker searches for the name of a particular fish species. Since the pause following the placeholder is very short (70ms), I analyse the placeholder as the first element (i.e. the modifying element) of a nominal compound (*ane bäre kofä* > *gastol*),¹⁴ rather than as a separate noun phrase (*ane bäre* > *kofä* > *gastol*). After a pause of 470ms, the speaker provides the name of the fish species *gastol* (cf. Figure 12), but the anaphoric demonstrative *ane* and the noun *kofä* 'fish' are not repeated (not: *ane gastol kofä*).

¹³Houses are built on posts with a sitting platform called *skiski* underneath.

¹⁴*Gastol* is the striped snakehead (*Channa striata*), which is an invasive species to the region. Hence, the word *gastol* is a loanword from Papuan Malay.

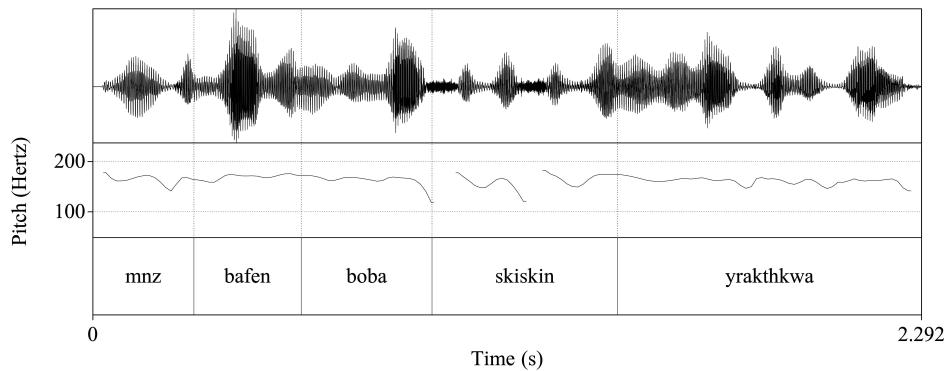


Figure 11: Audio analysis of [tci20111119-03 ABB 34]

- (16) komnzo *thu|kwthe/nzrm* ***ane bän̄e*** ... ***kofä fof*** ...
 only SG>3PL:PST:DUR\turn DEM PH (70ms) fish EMPH (470ms)
gastol
gastol(MA)
 'I was just turning those whatchamacallit fish ... (those) gastol (fish)'
 [tci20130903-03 MKW 125–126]

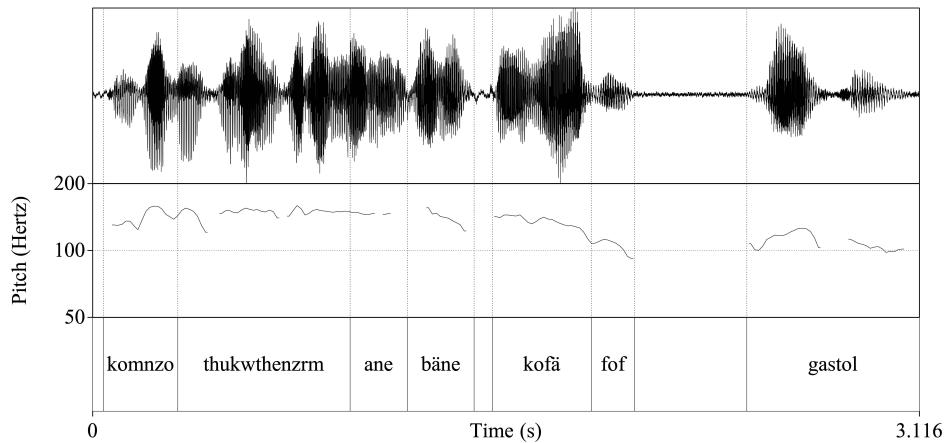


Figure 12: Audio analysis of [tci20130903-03 MKW 125–126]

Example (17) is clearer because all parts of the complex noun phrase are repeated (*ane bän̄e* > *ane trikasi*). At the same time, the example is interesting because the pause of 820ms (cf. Figure 13) comes after the anaphoric demonstrative,

i.e. in the middle of the delayed constituent. It is cross-linguistically common to have hesitation pauses after preposed function words (Himmelmann 2014: 935).

The structure of (17) is suggestive of another kind of analysis, namely a combination of the placeholder *bäne/baf* and a light verb placeholder ('be' in this case). The two clauses in (17) are hanging topic constructions that I translate with 'as for ...' (cf. Döhler 2018: 337), and in the corpus the speaker continues to explain how the story was passed down from the ancestors. A more suitable analysis might be that the entire topic construction acts as a placeholder substituting the following clause which is also a topic construction (*ane bäne mane rera > ane trikasi mane njatrikwé*).

- (17) *watik ane bäne mane re|r/a ane ... trikasi mane*
 then DEM PH which 3SG.F:PST:IPFV\be DEM (820ms) story which
n=ŋa\trik/wé *fof*
 IPST=1SG:NPST:IPFV\tell EMPH
 'Well, as for this whatchamacallit, as for this ... story that I've just told'
 [tci20131013-01 ABB 401–403]

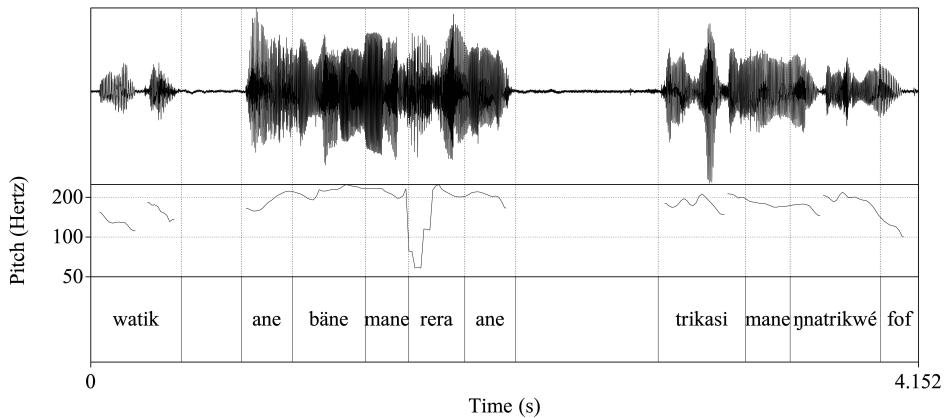


Figure 13: Audio analysis of [tci20131013-01 ABB 401–403]

In longer stretches of discourse, there is a framing structure in which the placeholder appears in an opening part, then there is some elaboration that reveals the substituted referent, and finally there is a closing part that repeats the structure of the opening part. An example of this is given in (18) below. The speaker talks about a man who has no children of his own, but he has adopted the children of

his wife, whose former husband died many years ago. The context here is one of taboo, because the speaker is in a taboo relationship with the deceased man, which means that he should not utter his name. We will see in §3.3 that this framing strategy is often used in taboo contexts.

- (18) *wati baf=ane=nzo nagayé ä|moneg/wr* ...
 then PH=POSS.SG=ONLY children 3SG>3PL:NPST:IPFV\look_after (2000ms)
mabata-a-fis kwark zöbthé mane ya|r/a, kabe fof
 PN-POSS-husband deceased first who 3SG.M:PST:IPFV\be man EMPH
 ... *nafane kabe ... masen ...* *wati nafane*
 (1650ms) 3SG.POSS man (600ms) PN (1450ms) then 3SG.POSS
nagayé=nzo ä|moneg/wr
 children=ONLY 3SG>3PL:NPST:IPFV\look_after
 ‘Well, he looks after whatsisname’s children ... Mabata’s late, first
 husband ... her man ... Well, he looks after only his children.’
 [tci20120814 ABB 217–221]

The repeated part may involve more than just the delayed constituent. In example (19), the copula clause is repeated several times by two speakers who are trying to find and negotiate the correct expression for ‘mixed’ or for ‘random order’. The example comes from a stimulus task, in which two speakers are asked to arrange a set of pictures into a story. In the final part of the task, the story is presented to a third participant. In (19), speaker RMA explains to the third participant that the set of pictures came in random order. He cannot think of the right word immediately and therefore uses a placeholder in a copula clause (*bäne thfrä ane*). He then corrects himself and uses an English insertion (*mix thfnrä ane*). Speaker TSA corrects RMA by using a Komnzo verb instead of the English insertion (*thafraksikaf thfrä ane*), and speaker RMA repeats this.¹⁵

- (19)
- RMA: *nzäthe zöbthé mane nzwan\ri/n* ***bäne***
 namesake first which 3SG>1DU:RPST:IPFV:VENIT\GIVE PH(ABS)
thf|rä/ ane ... mix thfn\rä/ ane
 3PL:RPST:IPFV\be DEM(.) mix(E) 3PL:RPST:IPFV:VENIT\be DEM
 ‘When (our) namesake first gave us the (pictures), they were
 whatchamacallit ... they came mixed.’ [tci20111004 RMA 305–307]

¹⁵Note that in (19), RMA’s response to TSA is not a perfect repetition, because there is a different proprietive case marker. The two variants, =*karä* and =*kaf*, however, do not differ in their semantics (Döhler 2018: 161ff.).

- TSA: *thafrak-si=kaf* *thf\rä/*
mix-NMLZ=PROP 3PL:RPST:IPFV\be
‘They were mixed.’ [tci20111004 TSA 206]
- RMA: *thafrak-si=karä* *thf\rä/*
mix-NMLZ=PROP 3PL:RPST:IPFV\be
‘They were mixed.’ [tci20111004 RMA 308]

The placeholder and the delayed constituent do not need to be adjacent. A frequent pattern in the corpus is that the placeholder remains in-situ, while the delayed constituent is postposed to the clause. One example is shown in (20) below. Note that there is a pause of 760ms between the clause and the delayed constituent (cf. Figure 14).¹⁶

- (20) *kofä mane baf=en kwa
fish(ABS) which PH=LOC 3PL:PST:DUR\enter (760ms) fish.trap=LOC
‘and the fish was going into the whatchamacallit ... into the fish trap.’
[tci20110802 ABB 64–65]*

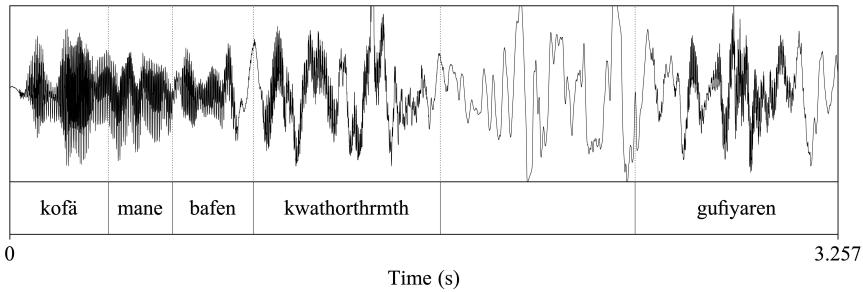


Figure 14: Audio analysis of [tci20110802 ABB 64–65]

3.3 Functions

Most of the examples of *bäne/baf* that we have seen so far involved cases of memory lapse or problems of accessing a lexical entry. It is those “tip-of-the-tongue” disfluency situations that are characteristic of placeholder uses. Hence, it is no surprise that many corpus examples also involve repair situations, as in (21). The speaker first produces a false start, the target of which is *mobilema*

¹⁶The audio quality of the example was not good enough to produce a meaningful pitch graph.

'because of the mobile phone'. Next she inserts a placeholder (*bänema*), and after a brief pause of 150ms (cf. Figure 15) she produces the delayed constituent: *radioma* 'because of the radio'.

- (21) *watik* -/mo/- ***bäne=ma*** ... *radio=ma* *noku=karä=nzo*
 then FS PH=CHAR (150ms) radio(E)=CHAR anger=PROP=ONLY
kwa\fark\wrmth
 3PL:PST:DUR\set_off
 'So they were leaving in anger because of the mo... **because of the**
whatchamacallit ... because of the radio.' [tci20131004-05 RNA 39]

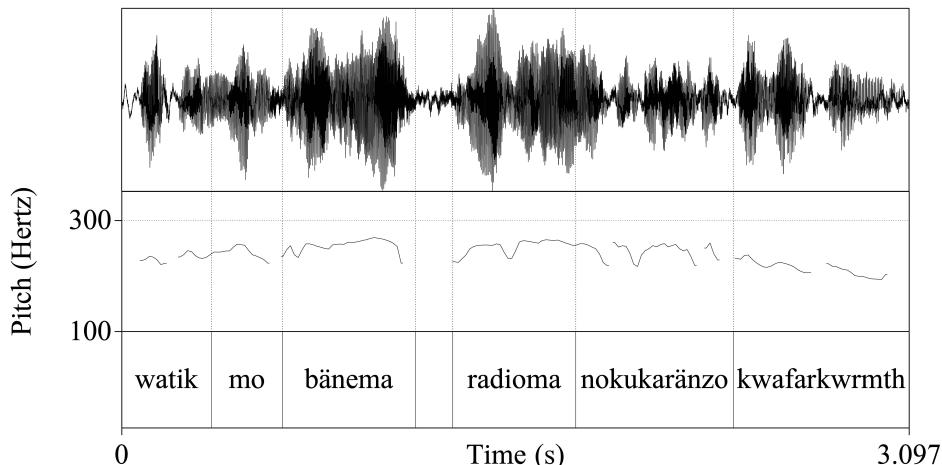


Figure 15: Audio analysis of [tci20131004-05 RNA 39]

The above example shows a context that is typical of placeholder uses, namely the occurrence of loanwords or ad-hoc insertions. This might be due to the fact that in some cases the speakers wish to express something for which there is no single word in the Komnzo lexicon, such as 'radio' (21), 'prison cell' (3) or 'Gastol fish' (16). Finding an ad-hoc insertion can lead to disfluency. An additional factor can be linguistic purism, in the sense that speakers try to avoid using words from another language. This is an important part of the linguistic ideology of Komnzo speakers. An example is (19) where the speaker is corrected by his interlocutor after using an ad-hoc insertion from English 'mix'.

Another context for using placeholders is to hint at certain cultural sensitivities such as avoidance, taboo, or face-saving. This has been reported in the literature for different languages such as Lao (Enfield 2003: 108ff.), Mandarin (Cheung

2015), Korean, and Japanese (Hayashi & Yoon 2006: 502ff.). For Komnzo speakers, such sensitivities may be triggered by topic (e.g. substance abuse or sexual relationships) or more commonly by certain kinship relations which dictate name avoidance. Especially for affinal kin, name avoidance is seen as a way of showing respect, and therefore one should not pronounce the personal names of certain individuals. It follows that name avoidance can clash with the need to identify a particular individual. Komnzo speakers solve this problem by employing tekonyms ('X's father' or 'X's husband') or by using the placeholder *bäne/baf*. While this strategy works well when interacting with people who share sufficient common ground, it is more difficult in recording situations that involve an outsider: a linguist fieldworker.

Consider example (22). The speaker talks about his sister, whose husband had died in an accident some 30 years ago. The speaker was in an avoidance relationship with the deceased husband, because he was his brother-in-law. On first mention, he uses a placeholder (*bafane mezü* 'whatsisname's widow'). After a pause of 1350ms, he pronounces the name (*Masenane mezü* 'Masen's widow'), and adds a tekronym (*albertaŋafe kwark* 'Albert's late father') for further elaboration. After another pause of 1580ms, he closes with a phrase that mirrors exactly the opening phrase (*nafaŋafane mezü* 'his father's widow'). The closing phrase is all the more peculiar, because he could have said simply *nafane name* 'his mother' (i.e. 'Albert's mother').

- (22) *mabata fi mezü zwa\m/nzrm ... baf=ane mezü*
PN 3SG.ABS widow 3SG.F:PST:DUR\stay (1350ms) PH=POSS.SG widow
re\r/a ... masen=ane mezü ... albert-a-ŋafe
3SG.F:PST:IPFV\be (370ms) PN=POSS.SG widow (1580ms) PN-POSS-father
kwark ... nafa-ŋaf=ane mezü
deceased (2300ms) 3SG.POSS-father=POSS.SG widow
'Mabata stayed as a widow. She was **whatsisname's widow** ... **Masen's**
widow ... **Albert's late father** ... **his father's widow**'. [tci20120814 ABB 39]

Examples (22) and (18) come from different sections of the same text. In both examples the speaker talks about his brother-in-law, and both examples show a striking similarity in that there is a kind of bracket structure with an opening and a closing part. The first mention of the brother-in-law is a placeholder in both examples, presumably signalling that this a taboo context.¹⁷ What follows is a careful elaboration during which the name to be avoided is in fact uttered. Finally,

¹⁷A similar context is found in Kalamang (Visser 2025 [this volume]) and Besemah (McDonnell & Billings 2025 [this volume]).

the bracket is closed with a phrase in (22) – or with a clause in (18) – that exactly mirrors the opening part. In Komnzo speech, this bracket structure is frequently found when talking about sensitive topics. It follows that the placeholder is used with a communicative goal rather than filling a disfluency. Figuratively speaking, the placeholder and the bracket structure set a stage on which is it permissible to break with the taboo.

Another use outside of disfluency is for managing turn-taking, more specifically for “gaining the floor”, which relates to conversational dynamics rather than conveying some communicative goal. Example (23) shows a short exchange from the picture task, in which speaker RMA comments on a picture card showing a man being dragged off by two police officers. The speaker TSA adds his own thoughts on the state of affairs, which RMA agrees with.

(23)

RMA: *aiwa ... frisman=é kabe
oh_no (500ms) policeman(E)=ERG.PL man(ABS)
y\thärku/nth
3DU>3SG.M:NPST:IPFV\drag*

‘Oh no, the two policemen are dragging away the man.’

[tci20111004 RMA 109–111]

TSA: *bäne=ma y\thärku/nth ηare mane
PH=CHAR 3DU>3SG.M:NPST:IPFV\drag woman(ABS) which
nz=ü\fn/zro
IPST=SG>3SG.F:NPST:IPFV:ANDAT\hit
‘That’s why they are dragging him away: It was the woman who he hit
just before.’*

[tci20111004 TSA 89]

RMA: *mh
INTERJECTION
‘mh (okay).’*

[tci20111004 RMA 112]

TSA’s successful interruption is achieved by starting his turn with the placeholder *bänema* ‘that’s why’ that has the entire following clause as its delayed constituent, as is indicated by the underlined font in (23). The initial placeholder creates a moment of anticipation that further elaboration is to come, thus, enabling TSA to take over the floor. There are no pauses in TSA’s turn, but there is a break in the intonation contour with a falling pitch on the last word of the first clause (*ythärkunth*) and a rising pitch on the first word of the second clause (*ηare*), which separate the two clauses prosodically (cf. Figure 16).

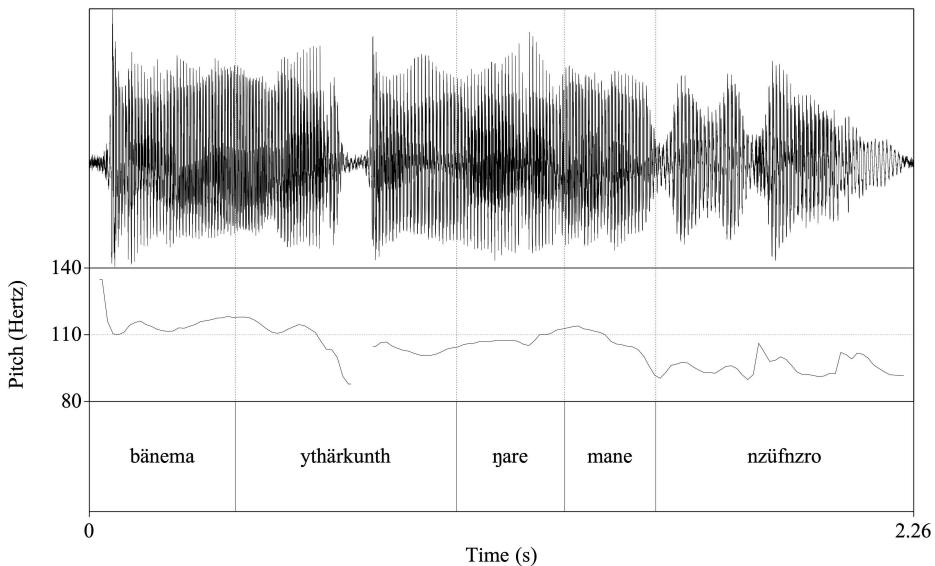


Figure 16: Audio analysis of [tci2011004 TSA 89]

3.4 Functional extensions

In most of the preceding examples, the placeholder mirrored the delayed constituent in its syntactic position and case marking. In this section, I argue that some of the inflected forms of *bäne* have widened their functional scope to include a non-placeholder function, namely they are used as clausal connectors for adverbial clauses.

Consider example (24), in which the speaker explains how he stores different species of yam in his storage house. The placeholder *bäne* is inflected with the purposive case (*bänemr*), which has a temporal meaning in this example ('until'). The placeholder functions as a connector of two otherwise independent clauses. There is no sign of disfluency in the example. The falling pitch on *bänemr* and subsequent reset of the pitch level separate the two clauses (cf. Figure 17). Thus, prosodically *bänemr* belongs to the first clause.

- (24) *ane fof e\mig/wre bæne=mr fobo kwa*
DEM EMPH 1PL>3PL:NPST:IPFV\hang PH=PURP DIST:ALL FUT
thra\rfik/wr
3PL:IRR:IPFV\grow
‘We hang them up until (the shoots) will grow from there.’

[tci20121001 ABB 24]

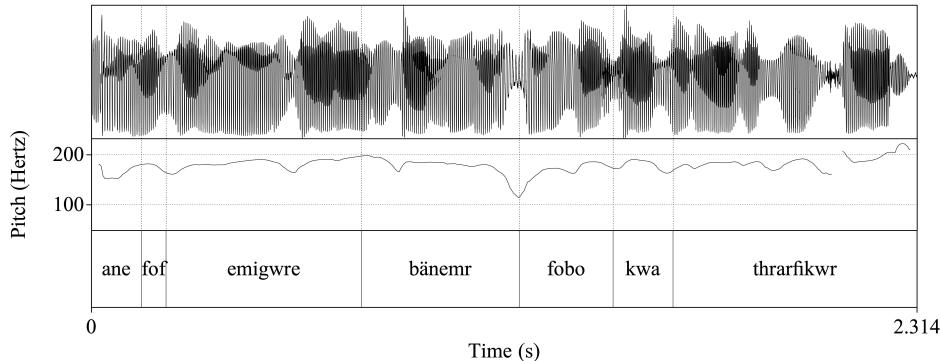


Figure 17: Audio analysis of [tci20121001 ABB 24]

Tokens of *bäne* such as (24) cannot be analysed as placeholders, nor can the following clause be analysed as a delayed constituent. Instead, they are part of a grammatical construction that connects two independent clauses, and it is the case marker on *bäne* that signals the semantic relation that holds between them. There are adverbials of reason (*bäne=ma* [PH=CHAR] ‘because’), manner (*bäne=me* [PH=INS] ‘thereby’), and purpose/time (*bäne=mr* [PH=PURP] ‘in order to’/‘until’). Note that Komnzo has additional strategies for adverbials including other types of connectors (e.g. *fthé* ‘when’, *monme* ‘how’) or nominalised verbs for non-clausal adverbials (Döhler 2018: 321ff.).

(25) is a second example of *bäne* as a clausal connector. This time it is inflected with the characteristic case, which I translate as ‘because’. Like in (24), *bänema* has a falling pitch and the following clause resets the pitch level (cf. Figure 18).

Note that there is a second token of *bäne* in the first clause of (25), inflected with the instrumental case (*bäneme*). This one ticks all the boxes for a placeholder: there is a short pause (after the verb) signalling a disfluency, the placeholder mirrors the delayed constituent in terms of case marking and syntactic position. Note that there is no falling pitch on *bäneme*.

- (25) *zöbthé bäne=me kwa w\rthaku/nzr ... zzarfa=me*
 first PH=INS FUT 3SG>3SG.F:NPST:IPFV\sprinkle (200ms) ginger=INS
bäne=ma gatha miyosé \rä/
 PH=CHAR bad taste 3SG.F:NPST:IPFV\be
 ‘First, he will sprinkle it with whatchamacallit ... with ginger, because it
 has a bad taste.’ [tci20130903-04 RNA 63–64]

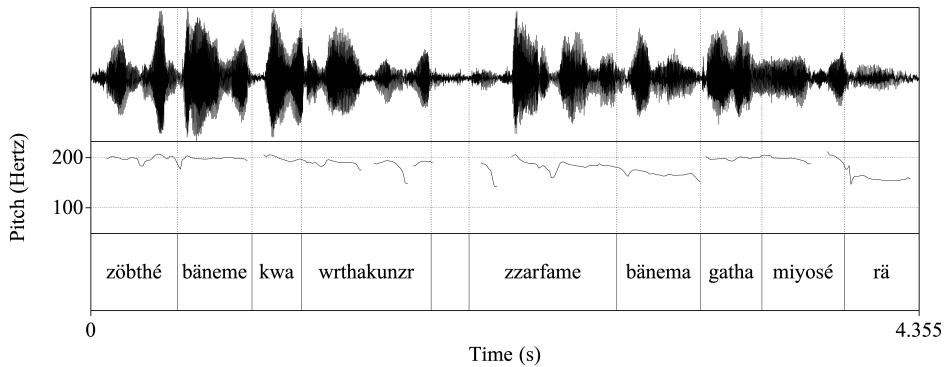


Figure 18: Audio analysis of [tci20130903-04 RNA 63–64]

Not all tokens of *bäne* in three inflections (CHAR, PURP, INS) are connectors, as we just saw with *bäneme* in (25) or with *bänema* in (21) and (23). Next to the meaning, it depends on syntactic and prosodic cues whether a specific example is best analysed as a placeholder or as a connector. Thus, it would be wrong to say that these inflections have grammaticalized to become connectors, but rather that they have widened their functional scope or their syntactic possibilities.

A reviewer of this chapter suggested that the development of the adverbial connector need not involve the additional step via a placeholder, but come directly from the medial demonstrative. Based on synchronic data, I cannot rule out this possibility. Moreover, Himmelmann (2014: 230) noted already that demonstratives in the “recognitional use” are often found as connectors of relative clauses, and there is an obvious link between recognitional deixis and placeholders (cf. Enfield 2003).

Nevertheless, I want to sketch out a scenario via the placeholder that seems to me more parsimonious. For this, let us think about the problem as bridging two ends of a spectrum: On the one hand, we have a clear placeholder use, which involves disfluency and some kind of mirroring of the delayed constituent. Most of the examples in this chapter fit this description. On the other end of the spectrum, we have examples like (24) and (25), where two clauses are connected, and where there is no disfluency.

One kind of bridging construction are cases in which the placeholder has the entire following clause as its delayed constituent, as in (23). A clearer example of this pattern is shown in example (26), in which the speaker introduces the protagonist of a story. He uses a placeholder in the first clause, and after a pause of 1100ms he informs us about the protagonist (cf. Figure 19).¹⁸ The function of

¹⁸Note that the characteristic case (=ma) covers both meanings of reason and aboutness (Döhler 2018: 157).

the placeholder in (23) and (26) is to create a certain anticipation that the speaker has more to say.

- (26) *trikasi bäné=ma kwa na\trik/wé ... kabe tnz yf*
 story PH=CHAR FUT 1SG>2SG:NPST:IPFV\tell (1100ms) man short name
sfrärm kukufia
 3SG.M:NPST:DUR\be PN
 'I will tell you a story about that one: The short man's name was
Kukufia.'
 [tci20100905 ABB 6–7]

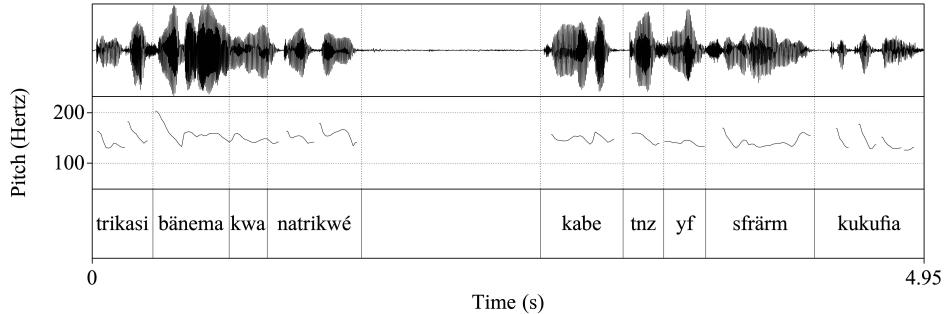


Figure 19: Audio analysis of [tci20100905 ABB 6–7]

All it takes for the second pillar of the bridge, is for the placeholder to occur after the verb, i.e. in final position. This happens frequently in a kind of afterthought expression which is introduced by *bäné*. In (27), the speaker explains the layout of yam tubers in his storage house. The placeholder (*bäné=mr*), flagged with the purposive case, can be translated as ‘in order to’ or ‘so that’, and its delayed constituent is the following clause, the afterthought. The pauses preceding (850ms) and following (300ms) the placeholder signal a disfluency (cf. Figure 20). Thus, the example has features from both ends of the spectrum: disfluency and clausal connector use.

- (27) *keke nya\fsi/nzre komnzo e\nak/wre ...*
 NEG 1PL:NPST:IPFV\count just 1PL>3PL:NPST:IPFV\put_down (850ms)
bäné=mr ... gb thra\rfik/wr zba
 PH=PURP (300ms) shoot 3PL:IRR:IPFV\grow PROX.ABL
 'We don't count (them). We just put them down ... so that ... the shoots
grow from here.'
 [tci20120805-01 ABB 33–35]

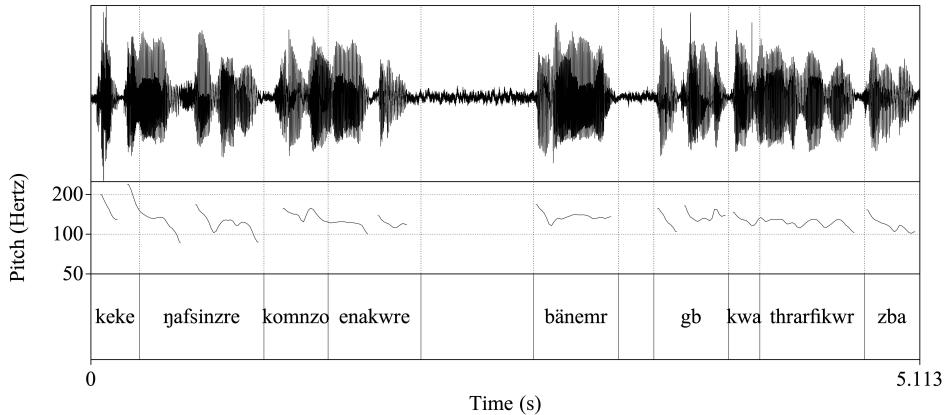


Figure 20: Audio analysis of [tci20120805-01 ABB 33–35]

I suggest here that afterthoughts like (27) provide a bridging context from which the function of *bäne* can be extended to include non-placeholder uses, i.e. to be used as different kinds of adverbial connectors. What drives this functional extension is the “hiatus moment” that is so typical of placeholders, i.e. the anticipation for further elaboration.

3.5 Multimodality

Multimodal aspects of placeholders have been largely neglected in the literature, but see the chapter on Northern Pastaza Kichwa (Rice 2025 [this volume]). An exception is Navarretta’s (2016) study on Danish hesitative fillers and simultaneously occurring gestures. In the examples from Komnzo, the placeholder *bäne/baf* is often accompanied by a hand gesture, often a pointing gesture. This does not come as a surprise, as the placeholder has developed from a demonstrative.

Note that there has not been a detailed analysis or description of gestures in Komnzo, nor is the corpus currently annotated for gestures. For these reasons, I can only give a rough estimate of the frequency of gesture co-occurrence. One can find gestures in about two thirds of the placeholder tokens. I include here three examples from the Komnzo data.

The first example comes from a conversation in the garden. Speaker STK refers to a road junction in the forest, but has problems finding the correct place name in his description. He uses a placeholder inflected with the allative case in (28). The accompanying gesture is a pointing gesture consisting of a short jerk of the left hand in the corresponding direction. This is highlighted in the still image with the red circle (cf. Figure 22). The audio analysis in Figure (21) shows the

time and length of the gesture with the grey overlay on the Praat picture. The gestural component is almost perfectly aligned with the placeholder, but not with the target word *Fothr Zfthfo*, which is produced after a short pause. Hence, the gesture functions like a support during the retrieval of the correct place name. In fact, the gesture alone establishes the correct spatial relationships by pointing in the corresponding direction.

- (28) *fä mane \rā/ bänē=fo ... fothr_zfth=fo*
 DIST which 3SG.F:N PST:IPFV\be PH=ALL (400ms) PLN=ALL
 ‘where (the road) turns to whatchamacallit ... to Fothr Zfth.’

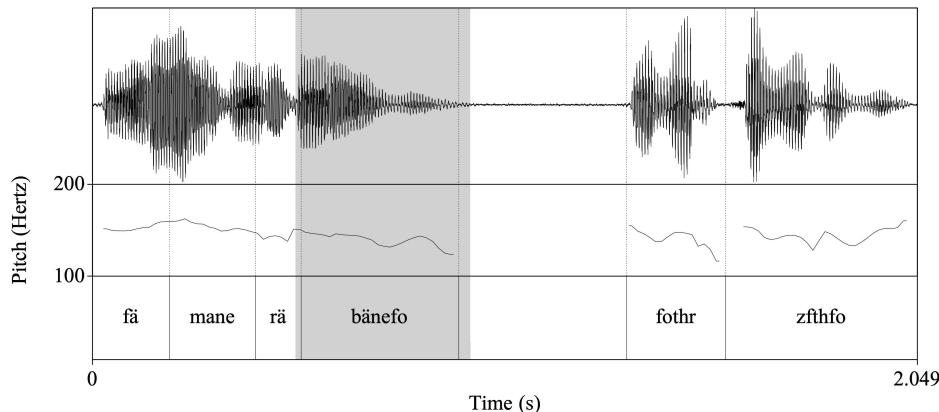


Figure 21: Audio analysis of [tci20130823-06 STK 143-144]

The second example comes from a conversational narrative in which speaker MAB talks about an event that happened a long time ago. His interlocutor CAM asks whether he was married at the time, to which MAB replies “No, I was just a boy”. To further emphasise his age, he adds that his beard had only just started to grow at the time. In (29), he uses a placeholder for the target word *fäk thäbu* ‘beard’. The accompanying gesture is that he strokes his left cheek with the fingers of his right hand (cf. Figure 24). The gestural component overlaps with more than just the placeholder. The audio analysis in Figure 23 shows that the gesture starts with the proximal demonstrative *zane*, lasts through the false start and the placeholder, and stops in the pause. The speaker does not produce the target word (*fäk thäbu* ‘beard’), because the gesture alone is sufficient to identify the referent.



Figure 22: Still image of [tci20130823–06 STK 143–144]

- (29) *komnzo kwa zane -/nzä/- bän̩e ... thf'rifik/wrm*
just FUT PROX FS PH (480ms) 3PL:PST:DUR\grow
'These watchamacallit were just about to start growing.'

[tci20130927-06 MAB 187]

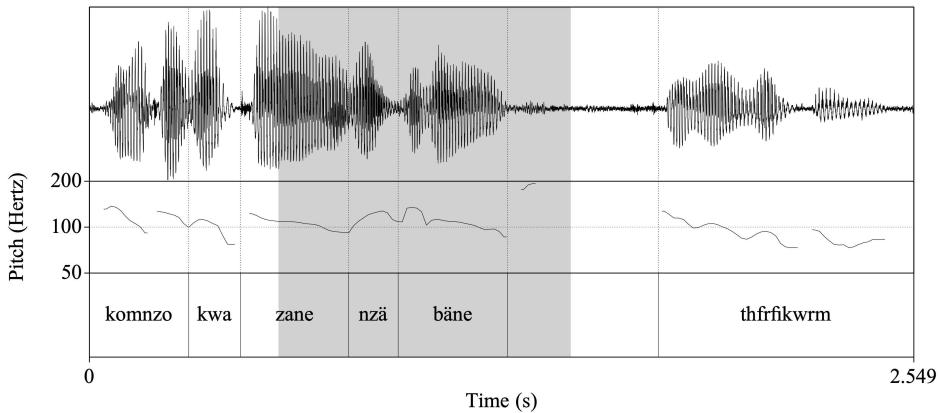


Figure 23: Audio analysis of [tci20130927-06 MAB 187]

In the third example, the gesture is more figurative. The example comes from the picture task, and it belongs to example (19) above: Speaker RMA explains to his



Figure 24: Still image of [tci20130927-06 MAB 187]

interlocutor, who is off the videoframe to the right, that the picture cards came in random order, i.e. they were mixed up. He has problems in lexical retrieval for this concept. First, he uses the placeholder, then an insertion from English (*mix*), and – after a prompt from the second participant TSA – a Komnzo word (*thafraksikarä*). The gesture consist of both hands spinning around each other on a horizontal left-to-right axis (cf. Figure 25), and he produces it two times. First, it occurs with the placeholder. Then he repeats the gesture with the English insertion, but not with the final expression in Komnzo. Here too, the gesture functions as a support channel during lexical retrieval.

The three examples presented here are anecdotal, but we can still derive some generalisations from them. First, gestures provide a parallel support channel during disfluencies. This is shown by the fact that the gesture only accompanies the placeholder and not the target word. Secondly, there is a broad range of gestures for such situations. The three examples presented here involve different manual gestures, but non-manual gestures such as lip- and head-pointing are also common. Thirdly, gestures vary in their function: They can be used to point to a target that is not present, as in the first example. They can be used to identify the target directly, as in the second example. Finally, they can be used to re-enact the target, as in the third example.



Figure 25: Still image of [tci20111004 RMA 305–307]

3.6 Frequency of *bäne/baf*

There has not been much mention in the literature on the frequency of place-holders. Podlesskaya (2010) reports 5 placeholders per 1000 words in a corpus of informal elicited narratives in Russian, and Zhao & Jurafsky (2005) report 6.68 placeholders per 1000 words in a conversational corpus of Mandarin. However preliminary these figures are, on first inspection the placeholder *bäne/baf* is much more frequent in the Komnzo data. There are 723 tokens in a corpus of 53,678 words, which amounts to 13.47 placeholders per 1000 words. This high figure has to be treated cautiously.

The figure of 13.47 per 1000 words is certainly too high, because some the 723 tokens are not placeholders, but connectors (cf. §3.4). For example, inflections such as *bäne=ma* [PH=CHAR] can be used as a placeholder, as in example (21), and as a connector, as in (25). The corpus has not been annotated in such a way that these can be searched for separately. To get an idea of the error rate, I have therefore excluded all of the inflections of *bäne* that can be used as connectors.¹⁹ This recount still results in a rather high figure of 9.71 placeholders per 1000 words. It follows that the true figure must lie somewhere between 9.71 and 13.47 per 1000 words.

¹⁹These are: *bäne=ma* [PH=CHAR] ‘because’, *bäne=mr* [PH=PURP] ‘in order to’, *bäne=me* [PH=INS] ‘thereby’). They add up to 202 tokens.

A reviewer of this chapter suggested that the high frequency may be an artefact introduced by the recording situation. The presence of an outsider linguist may exert pressure on the speakers to find the correct word and to avoid English insertions, which are permissible in natural conversation. I acknowledge that these pressures are real for Komnzo speakers as for anyone else. However, a brief examination of the data shows that this aspect has little influence in my data. To investigate this claim, I selected four more natural recordings. These are either purely observational recordings (even without my presence), or recordings in which I was only a spectator of an ongoing conversation. If we compare the frequency of *bäne/baf* in this subset with the overall frequency in the text corpus, we see that the frequency is only slightly lower: 13.27 placeholders per 1000 words.²⁰

A third explanation for the high frequency might be that *bäne/baf* is used in disfluency situations in which other languages employ hesitant fillers. While this is confirmed by my general impression of Komnzo speech, there is no way to measure relative frequency, as hesitant fillers are not consistently coded in the corpus.

4 Conclusion

I hope that the chapter helps to push forward the emerging typology of fillers. I close the chapter by summarizing a few interesting observations in Komnzo.

As I pointed out in §3.6, the figure of 13.47 placeholders per 1000 words is somewhat inflated. With a better coding of the data, this figure will come down, but not by much. The question of why Komnzo has a higher frequency than other languages will remain. Note that this observation extends to other languages in the Southern New Guinea region. For Bine, an unrelated Oriomo language spoken 200km to the East of Komnzo, I can report the staggering figure of 18.71 placeholders per 1000 words.²¹ For Evenki, an endangered Tungusic language spoken in Russia, China and Mongolia, Klyachko (2022) reports a token frequency of 12.6

²⁰The four recordings were: tci20130823-06 (a conversation in the garden between two speakers), tci20130927v-06 (a conversational narrative between two speakers), tci20130901-04 (a conversational narrative between three speakers), and tci20131004-05 (a conversation between six speakers). The combined figures for these four recordings were 70 placeholders out of 5274 words.

²¹629 tokens of the placeholder *nake* in a corpus of ca. 33,000 words. The coding problems are not an issue for Bine because (i) the placeholder *nake* is unrelated to demonstratives, and (ii) it is not used as clausal connector. That being said, the description of Bine is still in its infancy (cf. Döhler forthcoming).

placeholders per 1000 words²², and she explains this with a “lack of proficiency in some speakers” (2022: 213). This is not the case for Komnzo and Bine speakers. Although they are small languages – Komnzo is spoken by about 250 people, and Bine is estimated to have 2000 speakers – neither language is endangered, and the recorded speakers are fully competent. My general answer to the puzzle of the high frequency is that placeholders rather than hesitative fillers are the preferred strategy in disfluency situations.

The second point I want to raise here is the multi-functionality of *bäne/baf* in Komnzo. We have seen that its use goes well beyond that of a filler. It includes intentional uses with communicative goals such as signalling a taboo context. Moreover, it is used for interaction management in conversations. Lastly, in clause final position it has grammaticalized via an afterthought construction into a connector for adverbial clauses.

A third topic are co-occurring gestures, which open up a fascinating field of research for the study of fillers. As we have seen from the three examples in §3.5, Komnzo speakers use gestures as a visual support channel when they have problems retrieving the correct word from their mental lexicon.

Abbreviations

Abbreviations in the gloss line follow the Leipzig Glossing Rules. Additional conventions are given below:

...	pause	IPST	immediate past
(###ms)	measured pause in milliseconds	(MA)	loanword from Malay
ANDAT	andative ('away')	MED	medial
ANIM	animate	NPST	non-past
CHAR	characteristic	NSG	non-singular
(E)	loanword from English	ONLY	exclusive ('only X')
EMPH	emphatic	PH	placeholder
EPS	epistemic primacy	PLN	place name
FS	false start	PN	proper noun
HES	hesitator	PROP	proprietary ('having')
IC	inclusory case	PRIV	privative ('lacking')
INANIM	inanimate	RPST	recent past
		VENIT	venitive ('hither')

²²There are 350 tokens of the placeholder *ayi* in a corpus of about 27,700 words (Klyachko 2022: 213).

Acknowledgements

I thank the speakers of Komnzo for welcoming me to their village and taking on the task of teaching their language to me. The fieldwork was funded by the DOBES program of the Volkswagen Foundation and the Australian National University. I thank both of these institutions for their support of language documentation and description. Over the years, I have received much useful feedback during presentations that have touched on the topic of fillers in one way or another. Moreover, I owe thanks to the two anonymous reviewers of this chapter. Lastly, I thank Françoise Rose and Brigitte Pakendorf for their patience with the manuscript and for their useful feedback.

References

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Amiridze, Nino, Boyd H. Davis & Margaret MacLagan. 2010. *Fillers, pauses and placeholders*. Amsterdam: John Benjamins.
- Caballero, Gabriela & Alice C. Harris. 2012. A working typology of multiple exponence. In Ferenc Kiefer, Mária Ladányi & Péter Siptár (eds.), *Current issues in morphological theory: (Ir)regularity, analogy and frequency* (Current Issues in Linguistic Theory 322), 163–188. Amsterdam: John Benjamins. DOI: 10.1075/cilt.322.08cab.
- Carroll, Matthew J. 2016. *The Ngkolmpu Language with special reference to distributed exponence*. Canberra: Australian National University. (Doctoral dissertation).
- Cheung, Lawrence Yam-Leung. 2015. Uttering the unutterable with *wh*-placeholders. *Journal of East Asian Linguistics* 24(3). 271–308. DOI: 10.1007/s10831-014-9130-x.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/s0010-0277(02)00017-3.
- De Mulder, Walter. 1996. Demonstratives as locating expressions. In Martin Pütz & René Dirven (eds.), *The construal of space in language and thought*, 29–48. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110821611.29.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization* (Typological Studies in Language 42). Amsterdam: John Benjamins.
- Döhler, Christian. 2018. *A grammar of Komnzo*. Berlin: Language Science Press. DOI: 10.5281/zenodo.1477799.

- Döhler, Christian. 2023. Fluidity in argument indexing in Komnzo. *Open Linguistics (Lexical constraints on grammatical relations: Minority verb classes and restricted alternations)* 9(1). 20220201. DOI: 10.1515/opli-2022-0201.
- Döhler, Christian. Forthcoming. A sketch grammar of Bine. In Nicholas Evans & Sebastian Olcher Fedden (eds.), *The Oxford Guide to Papuan languages*. Oxford: Oxford University Press.
- Döhler, Christian. 2021. *Komnzo text corpus*. Zenodo. DOI: 10.5281/zenodo.4695271.
- Enfield, N. J. 2003. The definition of WHAT-d'you-call-it: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Evans, Nicholas. 2010. Semantic typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 504–533. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199281251.013.0024.
- Givón, Talmy. 2018. *On understanding grammar*. Revised edition [from 1979]. Amsterdam: John Benjamins.
- Givón, Talmy & Phil Young. 2002. Cooperation and interpersonal manipulation in the society of intimates. In Masayoshi Shibatani (ed.), *The grammar of causation and interpersonal manipulation* (Typological Studies in Language 48), 23–56. Amsterdam: John Benjamins. DOI: 10.1075/tsl.48.05giv.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders*, 33–66. Amsterdam: John Benjamins.
- Himmelmann, Nikolaus P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In Barbara Fox (ed.), *Studies in anaphora*, 205–254. Amsterdam: John Benjamins. DOI: 10.1075/tsl.33.08him.
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960. DOI: 10.1353/lan.2014.0105.
- Keenan, Edward L. & Stephen R. Anderson. 1985. Deixis. In Timothy Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon. Volume 3*, 259–307. Cambridge: Cambridge University Press.

- Klyachko, Elena. 2022. Functions of placeholder words in Evenki. In Andreas Hödlz & Thomas Payne (eds.), *Tungusic languages: Past and present* (Studies in Diversity Linguistics 32), 199–225. Berlin: Language Science Press. DOI: 10.5281/zenodo.7053369.
- Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67(3). 475–509. DOI: 10.2307/415035.
- Matthews, P H. 1974. *Morphology*. Cambridge: Cambridge University Press.
- McDonnell, Bradley & Blaine Billings. 2025. Choosing fillers in Besemah. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 169–209. Berlin: Language Science Press. DOI: ??.
- Mosel, Ulrike. 2012. Advances in the accountability of grammatical analysis and description by using regular expressions. *Language Documentation & Conservation Special Publication* 4. 235–250.
- Navarretta, Costanza. 2016. The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations. In *Linköping Electronic Conference Proceedings*, vol. 105, 55–61.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Rice, Alexander. 2025. *Mashti*: A multipurpose filler in Northern Pastaza Kichwa. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 373–439. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697595.
- Visser, Eline. 2025. Form and frequency of Kalamang fillers. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 251–273. Berlin: Language Science Press. DOI: ??.
- Zhao, Yuan & Dan Jurafsky. 2005. A preliminary study of Mandarin filled pauses. In *Proceedings of DiSS'05: Disfluency in spontaneous speech workshop*, 179–182. https://www.isca-speech.org/archive_open/diss_05/dis5_179.html.

Chapter 9

Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan)

Maïa Ponsonnet

Laboratoire Dynamique du Langage (CNRS & Université Lumière Lyon 2) &
University of Western Australia

This article analyzes the placeholder *keninjhbi* in Dalabon, a polysynthetic, non-configurational language of the Gunwinyguan family (non-Pama-Nyungan, northern Australia). Based on close to five hundred occurrences extracted from a 60-hour corpus, we observe that in line with the language's general architecture, *keninjhbi* is syntactically and morphologically flexible. It can replace verbal as well as nominal targets, and receive practically any of the morphology available to these targets. It can mirror the target's marking either integrally or partially; some occurrences even carry informative morphology that does not recur on the target.

Reflecting this flexibility and the range of options this placeholder offers, speakers vary in the ways they use *keninjhbi* – for instance in the word-classes they preferentially target, or in the amount of relevant information they package in the morphology allocated to the placeholder. The second part of the article compares speakers' respective tendencies in using the placeholder. The comparison demonstrates how in a polysynthetic language like Dalabon, where placeholders inherit the morphological agility of the language, speakers can adopt distinctive disfluency-management styles.

keywords: Dalabon, Australian languages, placeholders, polysynthesis, stylistics



1 Introduction

The volume in which this article is published takes an interest in “fillers”, defined as “non-silent markers of hesitation in disfluency” (Pakendorf & Rose 2025 [this volume]). Linguistic descriptions of fillers often assume, implicitly or explicitly, that since they reflect hesitation, fillers index speakers’ cognitive failures: Hayashi & Yoon (2006: 485), for instance, characterize them as words used “in contexts where speakers encounter trouble recalling a word or selecting the best word [...].” Linguists typically understand “disfluency” as a technical term stripped of prescriptive or evaluative connotations, and do not judge negatively what they may label speakers’ “trouble”. On the contrary, many linguists adopt a functional approach to disfluency, paying close attention to the broad range of pragmatic and discourse functions fillers fulfill beyond “disfluency repairs” (see for instance Keevallik (2010), or Kosmala & Crible (2022) for a review). Yet, beyond linguists’ relatively small networks, our choices of terms and definitions can also influence a less informed audience – or perhaps, in this case, simply encourage existing ideologies derived from prescriptive perspectives on language.

Indeed, in many day-to-day interactions, hesitations and “failures” can be judged negatively – whether this judgement is conscious or not – because they seem to index lesser proficiency (Fehringer & Fry 2007, de Jong 2016). As demonstrated by Bourdieu (2001) and many others, linguistic competence indexes and produces cultural capital and social power. Conversely, sounding linguistically incompetent incurs social costs. People across the world are well aware that they cannot escape this reality. As overt markers of disfluency, fillers are thus by nature caught up within a complex system of social signs, and we should expect speakers to handle them as such. They may try to avoid fillers altogether; or, on the contrary, recruit them to align or distance themselves from a norm, a group, or otherwise position themselves on a social chessboard (Silverstein 2003). Illustrations can be found in the public sphere, for instance in former French President François Hollande’s (2012–2017) use of the hesitative *euh*, and the perception of this strategy by his political adversaries¹ (see also Duez 1999; Boula de Mareüil et al. 2013). Of course, ordinary speech is rarely as thoroughly staged and planned as that of a presidential candidate. Yet, individual preferences in disfluency management are well attested, and inceptive stylistic manipulations of

¹As illustrated by the title of this article in *Le Point* during the 2012 campaign (22/01/2012): “L’UMP dégaine ‘le vrai discours de François euh... Hollande’”, “The UMP [Union for a People’s Movement] pulls out ‘the true speech of François er... Hollande’” – https://www.lepoint.fr/politique/l-ump-degaine-le-vrai-discours-de-francois-euh-hollande-22-01-2012-1422208_20.php#11 (I am not aware of any linguistic study on Hollande’s speech specifically).

fillers are detectable in their day-to-day usage (Laserna et al. 2014; Braun et al. 2023). In this article, I offer some considerations on how this plays out, and on which properties of fillers can lend themselves to such stylistic manipulations in Dalabon, a polysynthetic Australian language from the Gunwinyguan family.

As already mentioned, here I define fillers as “non-silent linguistic devices used in disfluencies”. Also in line with Pakendorf & Rose’s introduction to this special issue, among fillers I distinguish between hesitative interjections, which do “not occupy any specific syntactic slot within the structure of an unfolding utterance”; and placeholders, “referential expression[s] that [are] used as a substitute for a specific lexical item” and that hence “occup[y] a syntactic slot that would have been occupied by the target word” (Hayashi & Yoon (2006: 507, 490), respectively); see also Braun et al. (2023) for a taxonomy of disfluencies and filled pauses). This article examines placeholders, which, being syntactically “active”, can, in many languages, be morphologically enriched.

Dalabon, the language in focus here, is a member of the Gunwinyguan family (Australian, non-Pama-Nyungan), and is typical of this family in being polysynthetic and non-configurational. By definition, polysynthetic languages can aggregate a lot of morphology (Evans 2017). In such languages, a great deal of information is therefore conveyed by adding affixes and clitics to word stems; and in Dalabon, many of these morphological tools are highly flexible distributionally (Ponsonnet 2015). Dalabon has just one placeholder, but in line with the language’s architecture, it is syntactically and morphologically flexible: speakers can use it to replace words from most classes, and have at their disposal a large choice of items they can append to it. As a result, speakers’ preferred “styles” in using placeholders are easy to detect, even based on a relatively limited corpus. The language therefore offers an ideal arena to study variation in how different speakers use placeholders. So far, studies of placeholders in polysynthetic languages are rare (see Mithun 2025 [this volume] on Mohawk/Kanien’kéha’), and to my knowledge, none of them discusses the morphological flexibility of fillers in particular.

After presenting the Dalabon language and the speakers who contributed to this study in §2, in §3 I describe the syntactic and morphological behavior of the Dalabon placeholder, *keninjhbi*. In §4, I compare the distinctive tendencies or “styles” observed in the three speakers who contributed the most data to this study, and propose directions for further research in this respect.

2 Language and data

2.1 The Dalabon language

Dalabon is an Australian language from the Gunwinyguan family, one of the largest non-Pama-Nyungan families. Prior to colonization, Dalabon was spoken by a few hundred semi-nomadic hunter-gatherers in the western part of the Arnhem Land region in the Northern Territory (center north) of Australia. At the time of writing, Dalabon is severely endangered, numbering a small handful of fluent speakers. The descendants of Dalabon speakers live mostly in remote Indigenous communities to the east of the town of Katherine (Barunga, Beswick, Bulman, Weemol). They have adopted Kriol, an English-lexified creole spoken by thousands of speakers across the center north of Australia.

There exists no full grammar of Dalabon at this stage, but there is a dictionary (Evans et al. 2004), and a number of articles and theses describe various aspects of the language. This includes the verbal template (Evans & Merlan 2003; Ponsonnet 2014b: 61–64), morphological architecture (Evans 2017; Ponsonnet & Evans 2015), word classes including nominal subclasses and noun incorporation (Ponsonnet 2015), tense/aspect/mood categories (Evans & Merlan 2003), person prefixes (Evans et al. 2001), clause and argument marking (Evans 2006; Luk & Ponsonnet 2019; Ponsonnet 2021), noun incorporation and nominal subclasses (Ponsonnet 2015), demonstratives (Cutfield 2011), prosody (Evans et al. 2008; Ross 2011). Regarding semantics and lexical semantics, Bordluk et al. (2013) offers an inventory of the plant and animal lexicon, and Ponsonnet (2014b) examines the emotion domain.

Like many non-Pama-Nyungan Australian languages, and all Gunwinyguan languages, Dalabon is highly polysynthetic and agglutinative. The language is predominantly head-marking, with clausal arguments systematically cross-referenced by prefixes (and proclitics) on predicates. A discussion of the verbal template and a précis of Dalabon argument marking can be found in Ponsonnet (2014b: 60–62, 152–154). The full verbal template numbers 15 slots, although in ordinary speech, only a fraction of these are usually filled in a given utterance. In the first verb complex of example (1), only obligatory slots (person and mood prefixes, and TAM suffixes) are filled in. The second verb complex illustrates a richer morphology, with two aspectual markers and an incorporated adverb (*kakku-* ‘really’).

- (1) Dalabon (20120705b_004 118 [Film])

Bunu ka-h-na-ng, barra-h-dja-lng-kakku-yurd-minj.

3du 3sg/3du-R-see-PPFV 3du-R-FOC-SEQ-really-run-PPFV

‘She looked at the two of them, they were running fast.’

Dalabon noun phrases accommodate a set of case suffixes, as illustrated in (2) with the locative marker *-kah*. Noun incorporation, shown in (3) with *denge* ‘foot’, is frequent and in fact quasi-obligatory for certain classes of nouns like body-part nouns (Ponsonnet 2015).

- (2) Dalabon (20120720_003 193 [Film])
Oh klinik-kah bala-h-bobo-n ngale.
 INTJ clinic-LOC 3pl-R-go:REDUP-PRS INTJ
 ‘Oh, they’re going to the clinic, right.’
- (3) Dalabon (20110521a_002 030 [El])
Nga-h-dengu-bererde-mu.
 1sg-R-foot-ache-PRS
 ‘My foot aches.’

Apart from these generic typological properties, several other aspects of Dalabon grammar are worth keeping in mind to understand how placeholders pattern in the language. Overall, Dalabon exhibits a tendency towards non-configurationality (like most Australian languages, see Austin & Bresnan 1996) and morphological flexibility. Clause-level word order is syntactically free and pragmatically determined. Within the noun phrase, word order is also relatively free. As is the case in many Australian languages, Dalabon noun phrases can be discontinuous. This is illustrated in (4) where *lord-boyenj-ko* ‘big-bodied-pair’ and *yabbunh* ‘two’, both on the left of the verb, are part of the noun phrase headed by *biyi* ‘men’, on the other side of the verb. Morphology occurs either on the head or on the modifier. In (4), the dyadic marker *-ko* occurs on *lord-boyenj* ‘big-bodied’. In (5), the third-person possessive marker *-no* marks *wadda* ‘camp’, which is the possessed head of the noun phrase *ngal-Jane wadda-no-kah* ‘at Jane’s camp/place’. *Wadda-no* carries the morphological mark of its dependency to the verb *di* ‘stand’, of which it is a locative adjunct.

- (4) Dalabon (20111206a_001 38 [El])
Lord-boyenj-ko yabbunh barra-h-bobo-n biyi burrkunh.
 body-big-DYAD two 3du-R-go:REDUP:PRS man two
 ‘Two fat (big-bodied) men are walking by.’
- (5) Dalabon (20120708b_000 093 [Narr])
Ngal-Jane wadda-no-kah awudj buka-h-marnu-di.
 FEM-first.name camp-POSS-LOC house 3sg/3sg.h-BEN-stand:PRS
 ‘At Jane’s camp, stands her house.’

Generally, nominal morphology is distributionally flexible in Dalabon. That is, there is no categorical difference between the morphology nouns and adjectives allow. In addition, both these classes can receive some of the typically verbal affixes, for instance person prefixes marking the subject of a non-verbal predicate (see Evans et al. 2001 and Evans et al. 2004 for person prefix paradigms). In (6) for instance, the first person singular prefix *nga-* cross-references the subject of the nominal predicate *wurdurd* ‘child’ (noun), which is also inflected for tense.

- (6) Dalabon (30044/2007-4' [Narr])

Ngey	mah	wurdurd	nga-h-wurdurd-ninj,	budj	nga-h-dja-ni-nj.
1sg	CONJ	child	1sg-R-child-PIMP	bush	1sg-R-FOC-sit/be-PPFV
'As for me, as a child, when I was a child, I lived in the bush.'					

Conversely, typically nominal devices, such as possessive or case markers, can be appended to verbs. This is illustrated in (7), where *-burrng*, the standard possessive marker for third person dual (i.e. ‘their’, two people), is appended to the verb to cross-reference ‘them two’ as an oblique.² In (8), it is the locative *-kah*, which typically occurs on locative noun phrases, which marks the role of a subordinate verb complex.

- (7) Dalabon (20120713a_000 181 [Film])

Kenbo	kanh	ka-h-lng-dje-bruH-minj-burrng	djarrkno.
then	DEM	3sg-R-SEQ-nose-blow-PPFV-3du.POSS	together
'That's why she's upset with them then, with both of them.'			

- (8) Dalabon (20100716_002 [El])

Kenbo	bala-lng-mukka-ng	kanh	ka-ye-rakk-iyankah.
then	3pl-SEQ-cover-PPFV	DEM	3sg-SUB-fall-FUT-LOC
'Then they buried it, there where it would fall.'			

Overall, Dalabon morphology offers multi-purpose, flexible tools, and this has consequences on what speakers can do with placeholders.

2.2 Corpus and data set

The descriptions presented in this article draw on a 60-hour first-hand corpus collected between 2007 and 2012 in the First Nation communities called Weemol,

²This is the only way to cross-reference oblique arguments on verbs, but this cross-referencing is very rare, and therefore a marginal use of the set of pronouns commonly used to mark possessed nouns.

Beswick and Barunga, with eight speakers. All but one speaker were above sixty years old at the time, and only the youngest is still alive as I write this article. Five of them, including the three who produced the vast majority of the material, were female. These Dalabon consultants expressed pride in working towards documenting their language. They were eager to share their language with me, and to see it disseminated beyond the Dalabon community. In the analysis and “portrayal” of Dalabon I present here, I have done my very best to remain faithful to the spirit in which these speakers shared their language with me in the late 2000s and early 2010s. They usually preferred to be cited by name in writings where their data was used, as I have done in other publications. Due to the evaluative connotations surrounding the question of disfluency and the use of fillers (see §1), I make an exception here and use pseudonyms throughout the chapter. Family members and other Dalabon descendants are welcome to contact me, or my institutions if necessary, should they wish to access further details or the data.³

The corpus produced by these speakers is comprised of a range of audio and sometimes video recordings including narratives (mostly monological); free commenting on visual stimuli such as series of connected pictures, video clips, and thematically appropriate broad-audience films;⁴ as well as contextualized grammatical and semantic elicitation, particularly about body parts, bodily affects, and emotions (see Ponsonnet 2014a for a discussion). This covers a range of discourse genres, including relatively spontaneous speech. Although no lengthy conversations were recorded, some exchanges occasionally occurred between participants during group elicitation sessions, or with myself, and this is included in the corpus. The corpus is entirely transcribed in ELAN and therefore fully searchable.

Searching for instances of the placeholder returned a set of 476 occurrences. These were coded for a number of properties concerning the speaker, context of occurrence, target constituent (realization, word class, semantics), morphological enrichments on the placeholder and whether they matched those on the target, as well as presence or absence of a pause. The spreadsheet is available here: <https://doi.org/10.5281/zenodo.13304541>. The analyses below are entirely based on this data set; I have not carried out focused elicitation on placeholders.

³Personal address at the time of writing: maia.ponsonnet@cnrs.fr

CNRS team: <http://www.ddl.cnrs.fr/index.asp?Langue=EN>;

AIATSIS collections items: https://iats.ent.sirsidynix.net.au/client/en_AU/external/search/results?qu=ponsonnet&te=ILS; ELAR collection: <https://www.elararchive.org/dk0071/>

⁴In particular *Rabbit-Proof Fence* (Noyce 2003), *Ten Canoes* (de Heer & Djigirr 2006) and *Samson and Delilah* (Thornton 2009), all of which focus on Indigenous Australians’ lives (each in different historical times). Most speakers were already familiar with these movies before commenting on them for the purpose of linguistic documentation.

3 The Dalabon placeholder: *keninjhbi*

3.1 Form and functions

The most frequent form of the only Dalabon placeholder is *keninjhbi* [genij? bi]. Example (9) presents relatively typical occurrences where the form is fully articulated. Alternatively, speakers often contract it as *kenjhbi* [geŋ? bi] or *kenhbi* [gen? bi], or reduce it to *keninh* [genin?].⁵ A distinct word, *kenh* [gen?], which presumably results from a truncation of *keninjhbi*, has conventionalized as a self-correction interjection ('oops').

The etymology of *keninjhbi* is transparent: *keninh* is an interrogative pronoun, and *bi* is probably related to *biyi* [bi:], which in synchrony means 'person'. As flagged by Podlesskaya (2010: 12), interrogatives are a common etymological source for placeholders cross-linguistically. The variant *keninjhkun* [interrogative+genitive] is also used occasionally (see (11) below). The *keninjhkun* variant is about 10 times less frequent than *keninjhbi*, and exhibits no obvious difference in semantics or behavior. In this study all variants are included in the same data set and analyzed together.

- (9) Dalabon (20110518a_002 082 [Narr])

<i>Kanidjah bala-lng-bo-ninj</i>	<i>keninjhbi</i>	<i>-ngong</i>	[3.76s]	<u>kunj-<i>ngong</i></u>
DEM	3pl-SEQ-go-PIMP	PH-group		kangaroo-group
<i>bala-h-bo-ninj</i>	<i>keninjhbi</i>	<u><i>djakana...</i></u>		
3pl-R-go-PIMP	PH	bird.species		

'They were going there, all the whatsit... [pause] all the kangaroos were going there, the whatsit jacanas.'

Keninjhbi is a placeholder, not a hesitative interjection: it is syntactically integrated into the clauses where it occurs, temporarily or permanently replacing the "target constituent" it stands for. Syntactic integration is evident where *keninjhbi* carries morphology (as in the first occurrence in (9)). Where morphology is absent, prosody often indicates the syntactic role of *keninjhbi*, because it gives clues to the clausal delimitation and information structure. Morphology and prosody usually make it easy to identify the target of the placeholder in the rest of the utterance.

It is worth clarifying that in the vast majority of occurrences, *keninjhbi* is used to fill in for temporary disfluencies, irrespective of pragmatic motivations such as avoidance or euphemistic effects (as with the English *you-know-what* according

⁵The variants were also searched for and included in the data set.

to Enfield 2003). Enfield (2003) shows that in some languages, filler words can be recruited in contexts where pragmatic rules impose vagueness or indirectness (see Seraku 2024, among others). A pragmatic preference for indirectness has been documented in many Australian languages, including Bininj Gun-Wok, a close relative and neighbor of Dalabon (Garde 2008). Indeed, this preference applies in Dalabon too, however *keninjhbi* is not used to realize indirectness or euphemism.⁶ This is evident given that the target constituents of *keninjhbi* are realized more than 70% of the time. Furthermore, nearly half the occurrences in my data set feature some mark of hesitation, such as a pause or an explicit comment on the search (e.g. “ah, I can’t find this word”). In line with its filler role, *keninjhbi* frequently occurs in “lexically challenging” contexts, for instance when searching for relatively rare words like plant or animal species nouns (9), or before borrowings from Kriol (10), English, or other languages. When placeholders target borrowings, they most often represent words for objects or concepts that were not named in Dalabon, such as *peismeika* ‘pacemaker’ (10), *alikopta* ‘helicopter’, *djuwida* ‘sweater’, *eidyukeishen* ‘education’, *delebon* ‘telephone’, *bens* ‘fence’, *midjik* ‘music’, etc. While such words are not necessarily technical or rare, it is plausible that they may be somewhat less familiar to speakers, and/or less accessible when speaking Dalabon. The combination of occurrences in lexical elicitation of technical vocabulary, with those targeting natural species names, and occurrences involving borrowings, represent 40% of all occurrences of *keninjhbi* (and variants).

- (10) Dalabon (20120717_002 108 [ConvEl])
Ka-h-yidjnja-n keninjhbi peismeika.
 3sg/3sg-R-have-PRS PH pacemaker
 ‘He’s got a what-you-call-it pacemaker.’

3.2 Morphosyntactic properties

3.2.1 Syntactic versatility

In line with the tendencies observed in the rest of the grammar, Dalabon placeholders are syntactically flexible: they are attested with targets from a range of word classes. Perhaps reflecting the etymology of *keninjhbi*, noun targets (11) are most frequent, amounting to nearly 70% of all occurrences. This can include proper names, mostly of people and locations. Verb targets (12) are less frequent,

⁶My impressionistic assessment is that demonstratives are used for this purpose, but this remains a question for future research.

yet by no means rare, as they represent 30% of all occurrences. Predicative adjectives are also attested (13), and there is no evidence that any word class may *not* be targeted by *keninjhbi* (or in fact, any constituent, including phrasal constituents).

- (11) Dalabon (20110614_006 003 [Narr])
Bulu-ngan ka-h-wrokeb-m-inji langa keninjhkun...
father-1sg.POSS 3sg-R-work-VBLZR-PCUST LOC (Kr) PH
kenhbi-kah Bikûri-yad.
PH-LOC place.name
'My father used to work for **whatsit**... at **whatsit**, Pickery Yard.'
- (12) Dalabon (20120705b_001 011 [Stim])
Dohkardu munu kanh barra-h-djarrk... barra-h-djarrk-kenjhbi-mu
or.maybe LIM DEM 3du-R-together 3du-R-together-PH-PRS
barra-h-djarrk-yenjhyenjdju-ng yang.
3du-R-together-talk:REDUP-PRS language
'Or else the two of them together may just be... **the two of them whatyoucallit together**... the two of them are talking together.'
- (13) Dalabon (250909_89OK 0105 [El])
Ka-h-dorrung-kenjhbi ka-h-dorrung-mondi-duninj yuno pudiwan.
3sg-R-body-PH 3sg-R-body-good-INTENS CONJ pretty(Kr)
'She's **whatsit** in appearance... she's really good looking... you know, pretty.'

Podlesskaya (2010: 14) suggests that across languages, placeholders preferentially target nominals. Against this measure, the prevalence of noun targets in Dalabon appears less remarkable than the relative frequency of verb targets, which account for nearly one out of three occurrences. Overall, Dalabon *keninjhbi* can be characterized as a syntactically flexible placeholder.

3.2.2 Morphological versatility

Practically all morphological enrichments available to a target lexeme can be replicated on the placeholder that stands for this target. This is illustrated in (11) and (13) above for both nominal and verbal morphology.

There is no evidence of restrictions on the use of nominal morphology when *keninjhbi* targets a noun or an adjective. With verb targets, a notable exclusion is reduplication, but this could be due to the morphophonological make-up of

keninjhbi. Verbal reduplication is a relatively productive aspectual inflection for Dalabon verbs, with an iconic value around duration; however it is normally limited to one- or two-syllable verb roots.

When targeting verbs, the placeholder is assigned to its own conjugation class. Morphologically, many Dalabon verbs combine a “prebound” with a “thematic” (see Evans 2003: chap. 8) and Saulwick (2003) for comparable analyses on the neighboring Bininj Gun-Wok and Rembarrga languages). This is illustrated in (14), where the thematics appear in bold. Dalabon numbers about two dozen thematics. While each one tends to convey its own semantic “flavor”, they remain partly semantically opaque, or at least very versatile, and are therefore better analyzed as parts of the lexical verbs. They carry and determine TAM inflections, and therefore define the conjugation class to which each lexical verb belongs (Evans & Merlan 2003): compare (14a) to (14b), where *-mu* and *-minj* are respectively the present and past-perfective forms of the same thematic; and (14c) to (14f), which illustrate other thematics.⁷

(14) Dalabon (personal knowledge)

- a. *nga-h-boled-mu*
1sg-R-turn-THMC1:PRS
'I turn/change'
- b. *nga-h-boled-minj*
1sg-R-turn-THMC1:PPFV
'I turn/changed'
- c. *nga-h-buyh-won*
1sg/3sg-R-show-THMC2:PRS
'I show (it) to him/her'
- d. *nga-h-buyh-wong*
1sg/3sg-R-show-THMC2:PPFV
'I showed (it) to him/her'

⁷Strictly speaking, the morpheme partition should be *-m-u*, glossed -THMC-PRS, *-m-inj* glossed -THMC-PPFV, etc., but this would affect the readability of the glosses and add very little information. In this article, the neutral thematics *-mu* is glossed solely for its TAM value, i.e. *-mu* is glossed PRS, *-minj* is glossed PPVF, etc.

- e. *nga-h-dadj-ka*
1sg/3sg-R-cut-THMC3:PRS
'I cut it'
- f. *nga-h-ngadjdji-bun*
1sg-R-sneeze-THMC4:PRS
'I sneeze'

When *keninhbi* targets a verb, TAM inflections are always appended using the default, semantically neutral, typically intransitive thematic, the present form of which is *-mu* (14a and 14b). In other words, irrespective of the target, *keninhbi* is consistently assigned to the same conjugation paradigm, and affords the full set of TAM marking following this conjugation.

3.2.3 Morphological mirroring

The morphological flexibility of *keninhbi* enables it to "mirror" the morphology of its targets, i.e. the morphology of the target can be reproduced on the placeholder. Indeed, this is frequent, often with relatively complex morphological patterns. Mirroring is illustrated in (15) for a noun target, and in (16) for a verb target. Depending on speakers, full mirroring occurs on 30% to 50% of placeholders with nominal targets. Given that Dalabon verbal morphology is richer than nominal morphology, full mirroring concerns only 5% to 10% of occurrences of placeholders with verbal targets.

- (15) Dalabon (20100724_010 097 [Stim])

MT *djud-keninhbi-no ke*
neck-PH-3sg.POSS EMPH

MP *djud-kon-no* *kardu*
neck-fin-3g.POSS maybe

MT *nomo... djud-kon-no*
NEG (Kr) neck-fin-3sg.POSS

'MT Its dorsal-whatsit again...'

MP Its dorsal fin maybe?

MT No... Its dorsal fin.'

- (16) Dalabon (20110614_009 15 [Stim])

Nunda dja-h-dje-kenhbi-minj dja-h-dje-boled-minj.
DEM 2sg-R-nose-PH-PPFV 2sg-R-nose-turn-PPFV

'On this one you **whatsit-ed** your face... you turned your face.'

Mirroring can result in the placeholder carrying significant information. In example (12) for instance (repeated here for convenience), where full mirroring is realized, based on the placeholder the addressee already knows that the event at issue involves two participants in a harmonic classificatory relationship,⁸ with current, joint engagement. The retrieval of the target lexeme *yenjdjung* ‘speak’ adds that the participants are talking together, but much of the scene was already decipherable on the basis of the placeholder and its morphology, prior to lexical specification.

- (12) Dalabon (20120705b_001 011 [Stim])

Dohkardu munu kanh barra-h-djarrk... barra-h-djarrk-kenjhbi-mu
 or.maybe LIM DEM 3du-R-together 3du-R-together-PH-PRS
barra-h-djarrk-yenjhyenjdju-ng yang
 3du-R-together-talk:REDUP-PRS language

‘Or else the two of them together may just be... **the two of them**
whatyoucallit together... the two of them are talking together.’

Full mirroring can be complex, especially with verbal targets, and instead the mirroring is often partial: some of morphological elements are matched on the placeholder, others not. In (17) for instance, the placeholder carries the same person prefix as the target, and its suffix encodes the right TAM category (using the corresponding form from the placeholder conjugation, see (14) in §3.2.2). However, the target verb complex includes a benefactive applicative marker which is not mirrored on the placeholder.

- (17) Dalabon (20120720_003 170 [Film])

Buka-h-kenjhbi-minj bukah-marnu-yin-inj, baw kakkak...
 3sg/3sg.h-R-PH-PPFV 3sg/3sg.h-R-BEN-say/so-PPFV INTJ grandmother
 ‘She did whatsit... She told her, nah, grandmother..’

Placeholders with noun targets are often free of any morphology. This may in fact mirror the target if it has no morphology either, as in the second token in example (9) in §3.1 (*keninhbhi djakana* ‘whatsit *djakana*’, where *djakana* is a bird species). Sometimes it results in a mismatch, as in (18), where we see that the ergative marker *-yih* present on the target (*sista* ‘nun’) is absent on the placeholder.

⁸Two people either in the same generation or two generations apart, i.e. siblings, cousins, or grandparent and grandchild (see Evans et al. 2001).

- (18) Dalabon (20120721_000 159 [Film])

*Bala-h-yaw-djudju-mu. Kardu **kenhbi** sista-yih kardu.*

3pl-R-DIM-bath:REDUP-PRS maybe PH nun-ERG maybe

‘The little cuties are having a shower. Like the **whatsit**, the nun showers them maybe.’

Occasionally, a placeholder may carry informative morphology that does not appear on the target. In (19), the speaker explains how a needle is used to pierce seeds when making necklaces. The placeholder standing for *brikno* ‘needle’ carries an instrumental case marker *-yih*.⁹ This case is indeed semantically appropriate in the context, but is not repeated on the target. This “transfer” of morphology from the target to the placeholder aligns well with the Dalabon propensity to “morphological flexibility” (§2.1).

- (19) Dalabon (20100719a_003 021 [ConvEl])

*Yila-h-lng-dulubu-n kardu **keninjbi-yih** brikno*

1pl.excl/3sg-R-SEQ-pierce-PRS maybe PH-INST sharp

yila-h-dulubu-n kahnunh.

1pl.excl/3sg-R-SEQ-pierce-PRS DEM

‘We pierce it with a **whatsit**... we pierce it with a needle this one.’

Speakers differ in their tendencies to mirror more or less material from the targets onto placeholders, and this variation is discussed in §4.

Contrary to nouns, verbs in Dalabon always carry morphology: person prefixes and TAM suffixes are obligatory. Most frequently, both the person prefix and the TAM suffix are repeated, i.e. the pattern is “bilateral”. This results in the repetition of the entire verb complex, as illustrated above in (17), for instance. This pattern accounts for 97% of occurrences in the data set, yet two alternative patterns are observed. The first one is illustrated in (20), where the speaker repeats the TAM suffixes after the verb roots (future *-miyan* on the placeholder, future *-ngiyan* on the target), but does not repeat the prefixes (person prefix *ka-*, realis mood marker *h-*, sequential *lng-*). A last option, where no material is repeated at all and only the root is replaced, is illustrated in (21).¹⁰

- (20) Dalabon (2220909_70OK 0638 [El])

*Ka-h-lng-**keninjbi**-miyan [1.18 sec] yendju-ngiyan.*

3sg-R-SEQ-PH-FUT speak-FUT

‘He’ll **whatsit**... speak.’

⁹Instrumental and ergative are both expressed by the suffix *-yih*.

¹⁰In (21), the pause and prosodic reset after the placeholder, as well as the context, clearly indicate that *kenhbi* does not stand for an incorporated noun.

- (21) Dalabon (20110518b_002 125 [ConvEl])
Kenbo-yah dja-h-marnu-kenbi--, bimbu-yan.
 then-just 3sg/2sg-R-BEN-PH-- draw-FUT
 'She's going to whatsit draw it for you.'

These observations feed prior discussions on the delimitation of the prosodic word in Dalabon. In Gunwinyguan languages, verb complexes are typically analyzed as unitary words both from a grammatical and from a prosodic point of view – hence the representation of verb complexes as chains of morphemes separated by hyphens in the orthography (McKay 1975; Saulwick 2003; Evans 2003; Baker 2008; Kapitonov 2021). However, Evans et al. (2008) observed that verbal complexes are phonologically less coherent in Dalabon than in some other polysynthetic languages in the Gunwinyguan family (e.g. Bininj Gun-Wok). Dalabon complexes allow for pauses between verbal prefixes and the verb root. Along with some other prosodic and syllable-structure properties, this suggests incipient modifications in the structure of the prosodic words. The attestation of non-bilateral patterns of morphological repetition on placeholder targets, illustrated in (20) and (21) above, seems to confirm the same tendency. Of the three patterns presented above, the “bilateral” pattern is the only one that treats the verb complex as a block, while the others split it. In my corpus, this “bilateral” pattern is not only the most frequent by far, but also the only one used by the speaker who has Dalabon as her dominant language, vs Kriol (see §4.1). Increased familiarity with isolating languages like Kriol and English may influence the way speakers parse words in the polysynthetic language(s) in their repertoire.

4 Tendencies in use

Given the wealth of optional morphological enrichments and variation described above, Dalabon placeholders can be used in many different contexts and many different ways. Indeed, the three speakers who provided the most data to the corpus under examination seemed to adopt individual tendencies or “styles”. In this section, I attempt to characterize these styles based on quantitative comparisons between these three speakers. Naturally, given the nature of the corpus, the observations I propose cannot be conclusive. However, this case study suggests that important research on “disfluency styles” could be carried out based on synthetic, morphologically flexible languages numbering more speakers than Dalabon.

4.1 Use of placeholders does not correlate with proficiency

The three speakers who recorded the most data in the corpus – whom I will call S1, S2 and S3, see §2.2 – were all women above 60 years old at the time of recording. As I write, they are all deceased. Biographic information collected across a number of interviews revealed that each of them had a distinctive linguistic history, and Dalabon played a different role in their respective repertoires. A number of studies have shown that L2 speakers tend to use more fillers than L1 speakers (e.g. Fehringer & Fry 2007, de Jong 2016), yet others have shown that even for L2 learners, fillers endorse discourse functions beyond the management of hesitation (Kosmala & Crible 2022). With respect to the Dalabon speakers who provided data for this study, a first observation is that placeholder frequency does not correlate with lesser proficiency in Dalabon, at least not in any evident way.

While I am not well placed to assess proficiency as such, I can comment on the different “ranks” the Dalabon language used to occupy in each of these speakers’ respective repertoires. S1, who was about 10 years older than the other two, spoke only Dalabon during her early childhood. She had spent the first years of her life in the bush, away from white settlements and influences, and did not attend Western-style school. In her teenage and early adulthood, she acquired Mayali (another Gunwinyguan language),¹¹ Kriol, and English to a degree. During the years we worked together, Kriol was the language she used with her family and in most day-to-day contexts, yet as far as I can tell Dalabon remained her dominant language in terms of proficiency. She was regarded by other community members (as well as by linguists) as the primary authority for the Dalabon language.

S2 and S3, both born around 1950, belonged to the generation who “created” Kriol in their late childhood and teenage years (Ponsonnet 2010). S2 had learnt Dalabon first with her parents and grandparents, before adopting Kriol. She acquired English at a quasi-native level via school. Throughout her life, she continued to use Dalabon with some of her ascendant relatives for many years, and remained a relatively balanced bilingual, with a slight advantage for Kriol as her main day-to-day language.

S3’s linguistic biography was comparable to S2’s except that her parents were speakers of Kunwinjku (another Gunwinyguan variety¹²). She was therefore a balanced bilingual speaker of Kunwinjku and Kriol, and also acquired native-like English proficiency via school. She learnt Dalabon as an adult, in a deliberate endeavor to extend her multilingual repertoire to her grandparents’ endangered

¹¹A variety of Bininj Gun-Wok.

¹²Also a variety of Bininj Gun-Wok.

language. She positioned herself as a proud “new speaker” of Dalabon (Budrikis 2021). As far as I could judge, her Dalabon proficiency was not far from native-like, although she seemed to master less vocabulary than S1 and S2.

Extrapolating proficiency on the basis of linguistic biography, S1 was more proficient than S2, who was more proficient than S3. This aligns with their own self and mutual assessments, as well as local reputation as Dalabon speakers (and with my experience of these three speakers’ speech). Table 1 compares this observation with the frequency of placeholders per speaker across the corpus, where each of them contributed a similar proportion of narratives, stimuli-based elicitation, and lexical elicitation. As we can see, S3, who acquired Dalabon later in life, is not the one who used placeholders the most, since S2 uses them twice as often as S3 on average (1 every 2.5 min vs 1 every 5 min). The disparity in frequency of placeholders between S1 and S2, who are both native speakers of Dalabon, is considerable. Therefore, for these speakers and in this data, the frequency of placeholders does not correlate with how well they master the language. This lack of correlation does not mean that proficiency may not, in fact, play a role in how or how much speakers use placeholders. But it does show that with respect to the data set under examination, we need to look elsewhere to explain these three speakers’ tendencies. While the observation that using more fillers does not correlate with lesser proficiency may seem trivial, the question is worth further examination given the existence of “deficit” perspectives on fillers.

Table 1: Frequency of use by speaker

	Status of Dalabon	nb of tokens	frequency
S1	Dalabon as only L1	140	1/16 min
S2	balanced bilingual Kriol-Dalabon	259	1/2.5 min
S3	Dalabon as one of L2s	73	1/5 min

4.2 Differentiated styles

Apart from the clear difference in how frequently S1 and S2 used *keninjhbi*, in this corpus they generally used it in different contexts, targeting different parts of speech, and – consequently – with different morphological enrichments. In this section, I compare and characterize S1’s and S2’s respective styles in handling placeholders. S3’s style will be discussed briefly at the end of the section. Figures are recapitulated in Table 2, also at the end of the section.

A first, and key difference is in the parts of speech targeted by each speaker's placeholders. S1 and S2 both targeted nouns more often than verbs, but for S1 the difference was much greater: 75% nouns vs 16% verbs; while the distribution was 51% vs 40% for S2.¹³ This difference between S1 and S2 is statistically significant.¹⁴

These figures reflect that S1's placeholders occurred more frequently in contexts considered "lexically challenging" – i.e. in lexical elicitation, when discussing natural species, or with borrowings (see §3.1). Natural species form a very large lexical class with hundreds of items, some of them rarely evoked in these speakers' day-to-day lives. In (22), S1 comments on a movie where she identifies a specific plant. The corresponding noun does not occur elsewhere in my corpus, and S1 might not have heard it for several years prior to this recording session. Contexts comparable to this one, or to example (10) (in §3) with the borrowing *peismeika* 'pacemaker', represent 40% of S1's tokens, but only 12% of S2's (also a statistically significant difference¹⁵).

- (22) Dalabon (20120713a_000 072 [Film])
*Bula-h-beyu-ngiyan kanh **keninjhbi** kuladj.*
 3pl/3sg-R-fetch-FUT DEM PH spike-rush
 'They are getting this **whatyoucallit**, this spike-rush'

By contrast, in this data set, placeholders with verbal targets often stood for relatively common verbs. This is illustrated in (23), where S2's verbal target, *djawān*, is the standard verb for 'ask' and occurs at least two hundred times in the corpus. Overall, S1 used *keninjhbi* more often in contexts like (22), and S2 in contexts like (23). This correlates with the stronger prevalence of noun targets for S1, and the comparatively large proportion of verb targets for S2.

- (23) Dalabon (20110605_002 232 [Stim])
*Kardu **ka-h-kenjhbi-minj** malung, buka-h-ye-djawa-nj.*
 maybe 3sg/3sg-R-PH-PPFV first 3sg/3sg.h-R-COM-ask-PPFV
 'Maybe he **whatsit** first, he asked him'

In line with the type of contexts in which S1 and S2 used placeholders, on average S1 realized the targeted lexical item more often than S2. Overall, they both realized it relatively often: 77% of realizations for S1, and 69% for S2. The

¹³The figures do not add up to 100%, because they also account for other parts of speech, irretrievable targets, etc.

¹⁴Based on chi-square test returning $\chi^2(1) = 24.179, p < .001$

¹⁵ $\chi^2(1) = 40.659, p < .001$

difference between speakers is not very marked here (and indeed not statistically significant), but it converges with the speakers' preferred contexts. In a context like (22), preferred by S1, the form *kuladj* is in fact part of what the speaker is communicating about – i.e. teaching the addressee (myself) the name of the plant in question. In this situation, the placeholder helps hold the space while searching for a key lexical element. In a context like (23), on the other hand, eventually producing *djawān* for 'ask' is less relevant, as long as the addressee can grasp the intended meaning based on the morphology on the placeholder, and contextual clues. In such contexts, the speaker may care less about ultimately uttering the target.

Another difference between S1's and S2's placeholders is their morphological enrichments. Only 49% of S1's occurrences carry morphology, against 66% of S2's (another statistically significant difference).¹⁶ This, again, correlates with S2's preference for verb targets: these require some morphology, which placeholders with noun targets do not. This is reflected in the proportion of verbal versus nominal morphology in the speakers' respective data sets, as highlighted in Table 2. As a result, S2's tokens generally convey more information. This is firstly because they carry more morphology; and secondly because the morphology they carry, being verbal morphology, conveys crucial information about clause participants.

Bringing these observations together, we can portray S1's and S2's respective styles of placeholder management as follows. S1 used placeholders relatively "sparingly", privileging lexically challenging contexts, which typically involve nouns. Aiming for lexical accuracy, she took care to eventually realize the lexical target more than three quarters of the time. S2, by contrast, made more "liberal" use of placeholders, in the sense that she recruited *keninjhbi* more often, and in relatively unmarked contexts where the realization of the target mattered less. Relative to S1, S2 targeted more verbs, which were less lexically challenging. This in turn implied that S2's placeholders carried more morphological information, and this meant that realizing the target was even less relevant. Overall, S1's placeholders were geared towards lexical accuracy, while S2's seemed to privilege fluid communication.

This difference in orientation is evocative of Seyfeddinipur et al.'s (2008) distinction between using fillers for accuracy, i.e. to maintain accurate speech production; and using fillers for fluency, i.e. to avoid disruptions in the flow of speech. Their study examines these orientations in speech production. Here, the comparison between S1 and S2 suggests that a parallel distinction can be made at the level of discourse, i.e. with respect to speakers' motivation to use placeholders for better communication of content.

¹⁶ $\chi^2(1) = 11.132$, $p < .001$

These differentiated styles of “placeholding” may be interpreted in the light of the speakers’ respective proficiencies in Dalabon. Indeed, it may be the case that for S2, managing Dalabon fluency required slightly more cognitive effort than for S1, which could explain her introducing *keninjhbi* in place of frequent lexical items, including verbs. While this is very plausible, the comparison with S3 adds another perspective.

As mentioned above, S3 learnt Dalabon as an adult, yet she used *keninjhbi* less than S2 (see Table 1 in §4.1). Like S1’s, S3’s placeholders targeted many more nouns than verbs; but contrary to S1, this did not result from a preference for lexically challenging contexts, as only 5% of her tokens occurred in such contexts (vs 40% for S1 and 12% for S2). If we now compare S2 with S3, the higher rate of verb targets observed with S2 no longer seems to result solely from S2 being cognitively challenged when speaking Dalabon. Instead it could partly reflect a personal preference – and as pointed out above, this preference for verb targets allows her to make her placeholders morphologically richer, more informative, so that they optimize fluid communication.

Table 2: Distribution of features in speakers’ data

	S1	S2	S3
Placeholder frequency	1/16 min	1/2.5 min	1/5 min
Nominal targets	75%	51%	60%
Verbal targets	16%	40%	20%
Challenging contexts	40%	12%	5%
Target is realized	77%	69%	67%
PH has morphology	49%	66%	51%
of which verbal morphology	32%	59%	37%
of which nominal morphology	65%	41%	62%

Some clarification is in order here regarding the word “preference”. I am not suggesting that the styles sketched above involve a calculated, deliberate, or even conscious decision to use *keninjhbi* in one way or another. What I am proposing, instead, is that a speaker like S2 may be less reluctant to use placeholders in contexts where some other speakers may deploy more effort to avoid them. In other words, where S1 and S3 seemed to filter out placeholders with verb targets, S2 did not, and from there automatically exploited the resulting communication benefits that followed. Without involving a conscious decision, aiming to filter

out certain occurrences or not is a choice. In the case of placeholders particularly, such a choice is likely to be underpinned by a combination of contextual pressures (particularly in the context of linguistic elicitation), speakers' personal perceptions of linguistic norms, and general orientation in linguistic communication (see §1). All these questions call for investigation in future research in languages with larger populations of speakers.

5 Conclusions

The analysis of the morphosyntactic behavior and usage of placeholders in Dalabon I presented in this chapter has shown how disfluency-management devices can lend themselves to stylistic elaboration. This is particularly evident in a polysynthetic language like Dalabon, where rich morphology along with syntactic and morphological flexibility of the placeholder increase the range of options speakers can choose from when they use placeholders. Because the Dalabon placeholder, *keninjhbi*, can target virtually any word class, and endorse almost any morphology, different speakers use it in different contexts, load it with different morphology, and as a result recruit it for different communicative purposes. For many reasons – among which primarily the fact that it is no longer extensively spoken – Dalabon cannot really help us answer the questions that flow from the above observations, but I hope that they can be further explored in future research involving morphologically rich vital languages.

Acknowledgements

I am extremely grateful to all the Dalabon speakers who accepted to share their language with me and the broader scientific community. Most of the data analyzed here was collected thanks to a Individual Postdoctoral Fellowship (IPF0229) from the Endangered Languages Documentation Project.

Abbreviations

Glosses

Glosses not listed in the Leipzig Glossing Rules:

CONJ	conjunction	PPFV	past perfective
DIM	diminutive	R	realis
EMPH	emphatic	REDUP	reduplication
h	higher in animacy	SEQ	sequential
INTENS	intensifier	SUB	subordinate
INTJ	interjection	THMC	thematics
LIM	limitative	VBLZR	verbalizer
PCUST	customary past	x/y	argument x acting upon
PH	placeholder		argument y (e.g. 2sg/3sg)
PIMP	past imperfective	(Kr)	Kriol

Data types

[ConvEl]	conversation in elicitation
[El]	elicitation
[Film]	comment on movie
[Narr]	narrative
[Stim]	comment on visual stimulus

References

- Austin, Peter & Joan Bresnan. 1996. Non-configurationality in Australian aboriginal languages. *Natural Language & Linguistic Theory* 14(2). 215–268. DOI: [10.1007/BF00133684](https://doi.org/10.1007/BF00133684).
- Baker, Brett J. 2008. *Word structure in Ngalkgan*. Stanford: Center for the Study of Language & Information.
- Bordluk, Daisy, Nikipini Dalak, Maggie Tukumba, Lily Bennett, Rita Tingey Borodro, Margaret Katherine, Sarah Cutfield, Manuel Pamkal & Glenn Wightman. 2013. *Dalabon plants and animals. Aboriginal biocultural knowledge from southern Arnhem Land, north Australia*. Palmerston: Australia Department of Land Resource Management.
- Boula de Mareüil, Philippe, Gilles Adda, Martine Adda-Decker, Claude Barras, Benoît Habert & Patrick Paroubek. 2013. Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques [A quantitative study of discursive markers, disfluencies and speech overlaps in political interviews]. *TIPA. Travaux interdisciplinaires sur la parole et le langage* 29. DOI: [10.4000/tipa.830](https://doi.org/10.4000/tipa.830).
- Bourdieu, Pierre. 2001. *Langage et pouvoir symbolique*. Paris: Seuil.

- Braun, Angelika, Nathalie Elsässer & Lea Willems. 2023. Disfluencies revisited - are they speaker-specific? *Languages* 8(3). 155. DOI: 10.3390/languages8030155.
- Budrikis, Amy. 2021. *New speakers' perspectives on reinstating intergenerational language transmission of endangered indigenous languages in Western Australia*. Crawley: The University of Western Australia. (Doctoral dissertation).
- Cutfield, Sarah. 2011. *Demonstratives in Dalabon: A language of southwestern Arnhem Land*. Clayton: Monash University. (Doctoral dissertation).
- de Heer, Rolf & Peter Djigirr. 2006. *Ten canoes*. In collab. with Crusoe Kurddal, Jamie Gulpilil & Richard Birrinbirrin. https://en.wikipedia.org/wiki/Ten_Canoes.
- de Jong, Nivja H. 2016. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching* 54(2). 113–132. DOI: 10.1515/iral-2016-9993.
- Duez, Danielle. 1999. La fonction symbolique des pauses dans la parole de l'homme politique [The symbolic function of pauses in politician's speech]. *Faits de langues* 7(13). 91–97. DOI: 10.3406/flang.1999.1242.
- Enfield, N. J. 2003. The definition of WHAT-d'you-call-it: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Evans, Nicholas. 2003. *Bininj Gun-wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics, Research School of Pacific & Asian Studies, Australian National University.
- Evans, Nicholas. 2006. Who said polysynthetic languages avoid subordination? Multiple subordination strategies in Dalabon. *Australian Journal of Linguistics* 26(1). 31–58. DOI: 10.1080/07268600500531628.
- Evans, Nicholas. 2017. Polysynthesis in Northern Australia. In Michael Fortescue, Marianne Mithun & Nicholas Evans (eds.), *The Oxford handbook of polysynthesis*, 312–335. Oxford: Oxford University Press.
- Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 2001. Dalabon pronominal prefixes and the typology of syncretism: A Network Morphology analysis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 187–231. Dordrecht: Springer Netherlands.
- Evans, Nicholas, Janet Fletcher & Belinda Ross. 2008. Big words, small phrases: Mismatches between pause units and the polysynthetic word in Dalabon. *Linguistics* 46(1). 89–129. DOI: 10.1515/LING.2008.004.
- Evans, Nicholas & Francesca Merlan. 2003. Dalabon verb conjugations. In Nicholas Evans (ed.), *The non-Pama-Nyungan languages of northern Australia: Comparative studies of the continent's most linguistically complex region*, 269–

283. Canberra: Pacific Linguistics, Research School of Pacific & Asian Studies, Australian National University.
- Evans, Nicholas, Francesca Merlan & Maggie Tukumba. 2004. *A first dictionary of Dalabon (Ngalkbon)*. Maningrida: Maningrida Arts & Culture.
- Fehringer, Carol & Christina Fry. 2007. Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica* 41(1). 37–72. DOI: 10.1515/flin.41.1-2.37.
- Garde, Murray. 2008. Person reference, proper names and circumspection in Binjinj Kunwok conversation. In Ilana Mushin & Brett Baker (eds.), *Discourse and Grammar in Australian languages* (Studies in Language Companion Series), 203–232. Amsterdam: John Benjamins.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Kapitonov, Ivan. 2021. *A Grammar of Kunbarlang* (Mouton Grammar Library [MGL] 89). Berlin: De Gruyter Mouton. DOI: 10.1515/9783110747058.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Kosmala, Loulou & Ludivine Crible. 2022. The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1). 216–239. DOI: 10.1177/00238309211010862.
- Laserna, Charlyn M., Yi-Tai Seih & James W. Pennebaker. 2014. Um... Who like says you know: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology* 33(3). 328–338. DOI: 10.1177/0261927X14526993.
- Luk, Ellison & Maïa Ponsonnet. 2019. Discourse and pragmatic functions of the Dalabon ‘ergative’ case-marker. *Australian Journal of Linguistics* 39(3). 287–328. DOI: 10.1080/07268602.2019.1623758.
- McKay, Graham. 1975. *Rembarrnga: A language of Central Arnhem Land*. Canberra: The Australian National University. (Doctoral dissertation).
- Mithun, Marianne. 2025. Fillers and beyond. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 341–371. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697593.
- Noyce, Phillip. 2003. *Rabbit-proof fence*. In collab. with Doris Pilkington, Christine Olsen, Everlyn Sampi, Tianna Sansbury & Kenneth Branagh. [http://en.wikipedia.org/wiki/Rabbit-Proof_Fence_\(film\)](http://en.wikipedia.org/wiki/Rabbit-Proof_Fence_(film)).

- Pakendorf, Brigitte & Françoise Rose. 2025. Fillers in the world's languages: A refined typology. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 1–39. Berlin: Language Science Press. DOI: 10.5281/zendodo.15697575.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Ponsonnet, Maïa. 2014a. Documenting the language of emotions in Dalabon (Northern Australia). Caveats, solutions and benefits. In Aicha Belkadi, Kakia Chatsiou & Kirsty Rowan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory 4*, 1–13. London: School of Oriental & African Studies, University of London.
- Ponsonnet, Maïa. 2010. “Brainwash from English”? Barunga Kriol speakers’ views on their own language. *Anthropological Linguistics* 52(2). 160–183.
- Ponsonnet, Maïa. 2014b. *The language of emotions: The case of Dalabon (Australia)* (Cognitive Linguistic Studies in Cultural Contexts 4). Amsterdam: John Benjamins.
- Ponsonnet, Maïa. 2015. Nominal subclasses in Dalabon (South-western Arnhem Land). *Australian Journal of Linguistics* 35(1). 1–52. DOI: 10.1080/07268602.2015.976900.
- Ponsonnet, Maïa. 2021. Transfer and applicative constructions in Gunwinyguan languages (non-Pama-Nyungan, Australia). In Myriam Bouvieret (ed.), *Give constructions across languages* (Constructional Approaches to Language 29), 121–143. Amsterdam: John Benjamins.
- Ponsonnet, Maïa & Nicholas Evans. 2015. Emotional valuation in Dalabon diminutives. In Nicola Grandi & Livia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 401–407. Edinburgh: Edinburgh University Press.
- Ross, Belinda Britt. 2011. *Prosody and grammar in Dalabon and Kayardild*. Parkville: The University of Melbourne. (Doctoral dissertation).
- Saulwick, Adam. 2003. *Aspects of the verb in Rembarrnga, a polysynthetic language of northern Australia: Grammatical description, texts and dictionary*. Parkville: University of Melbourne. (Doctoral dissertation).
- Seraku, Tohru. 2024. Placeholders in crosslinguistic perspective: Abilities, preferences, and usage motives. *Linguistics* 62(6). 1473–1498. DOI: 10.1515/ling-2023-0068.

- Seyfeddinipur, Mandana, Sotaro Kita & Peter Indefrey. 2008. How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition* 108(3). 837–842. DOI: 10.1016/j.cognition.2008.05.004.
- Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication. Words and Beyond: Linguistic and Semiotic Studies of Sociocultural Order* 23(3). 193–229. DOI: 10.1016/S0271-5309(03)00013-2.
- Thornton, Warwick. 2009. *Samson and Delilah*. In collab. with Warwick Thornton, Beck Cole, Rowan McNamara, Marissa Gibson & Mitjili Napanangka Gibson. [http://en.wikipedia.org/wiki/Samson_and_Delilah_\(2009_film\)](http://en.wikipedia.org/wiki/Samson_and_Delilah_(2009_film)).

Chapter 10

Fillers and beyond

Marianne Mithun

University of California, Santa Barbara

In important work, Hayashi & Yoon (2006) distinguish types of fillers, elements used “in contexts where speakers encounter trouble recalling a word or selecting the best word” (2006: 485). Their placeholders are referential expressions used as a substitute for a specific lexical item. Their interjective hesitators are used to merely delay the production of the next element. Both roles are often filled by demonstratives, but while placeholders are produced as a syntactic constituent of a sentence, hesitators are not (2006: 489). Hayashi & Yoon (2006) propose that demonstratives functioning as interjective hesitators have been “pragmaticized for the discourse function of monitoring the flow of the production of an utterance to such an extent that they have diverged from ordinary demonstratives in terms of syntactic distribution, and referentiality” (2006: 488). Their examples are drawn from a range of languages, with special focus on Japanese, Korean, and Mandarin. Striking parallels can be seen in a typologically quite different language in another part of the world. In Mohawk, an Iroquoian language indigenous to North America, demonstratives are used similarly not only as placeholders and interjective hesitators, but also as devices for structuring the flow of information.

keywords: demonstratives, prosodic structure, Mohawk, complement constructions, relative clause constructions

1 Introduction

Fillers have usually been missing from language documentation, in good part because speakers do not typically consider them integral elements of the language; in fact, they are rarely even aware of them. Before the easy accessibility of audio or video recording technology, it was nearly impossible to document their



natural use; they would rarely make it into the record when speakers were creating written materials, translating, or dictating word by word for transcription. Advances in recording technology, however, have now made such documentation possible, and the results are revealing aspects of language processing which might fill in some detail in our understanding of certain pathways by which discourse patterns and even grammatical constructions can emerge.

Here structures involving a common type of filler, demonstratives used as placeholders, are described for Mohawk, a language of the Iroquoian family indigenous to northeastern North America. The same pattern is used on occasion for interjective hesitators. It occurs much more pervasively, however, in contexts not triggered by disfluency. The development of the pattern can be attributed to cognitive factors underlying the processing of information. In what follows, §2 describes the general structure of Mohawk, §3 the demonstratives, §4 their use as placeholders, §5 their use as hesitators, §6 their use in a conventionalized discourse structure, and §7 their appearance in contexts where complex syntactic structures are used in other languages.

2 Mohawk

Mohawk or Kanien'kéha' (MOH) is a language of the Northern branch of the Iroquoian family, spoken in six main communities in modern Quebec, New York State, and Ontario. Material cited here is drawn from a corpus of approximately 70,000 words of unscripted speech in a variety of genres, both conversation and monologue, from over 80 speakers. All six communities are represented.

The language is typologically polysynthetic, with potentially elaborate templatic verb morphology and extensive noun incorporation. It is head-marking: core arguments are represented by pronominal prefixes in the verb, one argument for intransitives and two for transitives. Nouns are unmarked for case, gender, or number. Alignment follows a basically agent/patient pattern. Constituent order is primarily pragmatically determined: constituents are arranged essentially in descending order of newsworthiness.

Three lexical categories are distinguished by their morphological structure: particles, nouns, and verbs. Particles by definition have no internal structure, though they may be compounded. They serve a variety of discourse, syntactic, and adverbial functions. Morphological nouns serve as referring expressions. Morphological verbs can function not only as predicates but also as referring expressions, often without any overt nominalization, and as clauses and even sentences on their own. Kinship terms, a subset of verbs more often used refer-

entially, are sometimes classified as an additional category. Verbs are noticeably more pervasive in speech than in many other languages.

Some of these structures can be seen in example (1). Transcription is in the community orthography. Digraphs <en> and <on> represent nasalized vowels [ʌ] and [ɥ], and <i> before a vowel represents a glide [j]. The layout here and throughout reflects prosody: each line represents a separate intonation unit (prosodic phrase). Intonation units are characterized by a coherent prosodic contour, typically beginning with a full or partial pitch reset, and often but not always preceded by a pause. In the transcription here, commas indicate non-final intonation unit boundaries, periods terminal prosodic boundaries, and dashes <--> truncated units. All examples are drawn from unscripted speech.

(1) Mohawk polysynthesis and head marking

VERB

Iahonwaia' ténhawe',
 i-a-honwa-ia't-enhaw-e'
 TRL-FAC-M.PL>M.SG-body-take-PFV
 they physically took him away
 'They took him away,

NOUN

karihtòn:ke,
 ka-rihton=’ke
 n-police=place
 police place

PARTICLE VERB

tánōn' wahonwahnhó:ton'.
 tanon' wa-honwa-nho-ton'
 and FAC-M.PL>M.SG-opening-cover-PFV
 and they door closed him
 and put him in jail.

PARTICLE PARTICLE VERB

Ó:nen	kati'	iontate'kèn:'a	NOUN VERB
onen	kati'	iontate-'ken'=a	Wá:ri iontátiaits,
then	in.fact	FI>FI-have.as.ygr.sibling=DIM NAME	Wari iontat-iat-s
then	in fact	FI>FI-have.as.ygr.sibling	Mary one calls her
Their sister Mary			

VERB

taionkwatewennátaħse'
t-a-ionkwa-ate-wenn-at-ahs-e'
CSL-FAC-FI>1SG-MID-voice-insert-BEN-PFV
she her voice inserted to me
phoned me.'

[Watshenní:ne' Sawyer, speaker]

3 Mohawk demonstratives

Mohawk contains two basic demonstratives: a proximal *kí:ken* ‘this, these’, and distal *thí:ken* ‘that, those’, as well as a discourse demonstrative *né*; which refers either anaphorically or cataphorically to something mentioned in the discourse. Examples are in (2), (3), and (4).

- (2) Proximal demonstrative

Iakwaterohrókha' *kí:ken.*
iakwa-ate-rohrok-ha' kiken
1EXCL.PL.AGT-MID-watch-HAB this
'We would watch this.'

[Watshenní:ne' Sawyer, speaker]

- (3) Distal demonstrative

Takónhewe' *thí:ken.*
ta-k-onhew-e' thiken
CSL.FAC-1SG.AGT-sweep-PFV that
'I swept that up.'

[Kahentoréhtha' Marie Cross, speaker]

- (4) Anaphoric discourse demonstrative

[Following description of how to make traditional cornbread]

Né:' eniákwake' *ohrhon'kè:ne.*
ne:=e en-iakwa-k-e' o-rhon'ke-hne
that=is FUT-1EXCL.PL.AGT-eat-PFV N-dawn-place
that we will eat it morning
'That is what we would eat in the morning.'

[Watshenní:ne' Sawyer, speaker]

The first two demonstratives can occur in apposition with a nominal, as in (5) and (6). Note that the demonstratives and nouns here were spoken in separate intonation units.

(5) Appositive proximal demonstrative

Sok ki' kí:ken,
then so this one
'So then this one,

o-- okwáho,
o-kwáho,
N-wolf
the wolf,

<i>wahatakia'takéhnha'</i>	<i>wahanòn:take'.</i>
wa-ha-at-ia't-akenha'	wa-ha-non't-a-k-e'
FAC-M.SG.AGT-MID-body-help-PFV	FAC-M.SG.AGT-milk-LK-consume-PFV
he helped himself	he consumed milk
helped himself to milk.'	

[Nancy Jackson, speaker]

(6) Appositive distal demonstrative

Tsi nòn:wa' takhará:tate' *thí:ken,*
tsi nonhwa' ta-k-haratat-e' thiken
at now CSL.FAC-1SG.AGT-lift-PFV that
'And then I lifted that thing,

ne kaiarahrónnion'.
ne ka-iar-a-hronnion-'
ART N-bag-LK-be.on-ST
the mattress.'

[Marie Kahentoréhtha' Cross, speaker]

These two demonstratives also occur, though more rarely, within the same intonation unit as a following nominal.

(7) Prosodic integration

<i>l:se' ken sá:wén</i>	<i>kí:ken truck?</i>
<i>ise' ken sa-wén</i>	<i>kiken truck</i>
2 Q 2SG.AL.POSS-possession this	truck
'Is this truck yours?'	

[Watshenní:ne' Sawyer, speaker]

(8) Prosodic integration

<i>Ioterihzionhátie'</i>	<i>thí:ken o'tá:ra'.</i>
<i>io-ate-rihs-i-on-hatíe'</i>	<i>thiken o-'tar-a'</i>
N.PAT-MID-disappear-ST-PROG that	N-clan-NS
'Those clans are disappearing.'	

[Watshenní:ne' Sawyer, speaker]

The discourse demonstrative *né:* does not occur with a nominal in this way. An article *ne* has developed from it, however, which does precede a referring expression. The article is short and unstressed and can fuse as a proclitic to a following vowel-initial nominal. Unlike its English counterpart, it also occurs with possessed nominals and proper names. Traditionally it has meant 'the aforementioned', but it is coming to be used as a more general definite article. Both the anaphoric demonstrative *né:* and the article *ne* can be seen in (9), as well as fusion of the article with the following vowel-initial first person pronoun: *ne i:'i > ní:*.

(9) Discourse demonstrative *né:* and article *ne*

Tóta.

'Gramma.'

<i>Né:' ní:' iakhina'tónhkhá'</i>	<i>ne Gramma.</i>
<i>ne:=e ne=i'i iakhina'tonhkw-ha'</i>	<i>ne Gramma</i>
that=is ART=1 1EXCL.PL>FI-call.by.name-HAB ART Gramma	

'That is what we called Gramma.'

[Watshenní:ne' Sawyer, speaker]

The proximal and distal demonstratives can co-occur with the article.

(10) Demonstrative combination

<i>Ok ò:ni' sénhá' kí:ken ne thotiièn:sa.</i>
ok ohni' senha' kiken ne t-hotii-en'sa
and also more this ART CSL-M.PL.PAT-be.young
'And it's also more these young people.'

[Joe Awenhráthon Deer, speaker]

All three of the demonstratives, *kí:ken*, *thí:ken*, and *né*: are morphological particles with no internal structure, though the first two can be traced historically to sequences *ken’ i:ken* ‘here it.is’ and *tho i:ken* ‘there it.is’. The modern forms are often shortened to *ki:* and *thi:* respectively. They can refer to persons, animals, objects, places, situations, ideas, etc., but, like nouns, they do not distinguish gender, number, person, or grammatical role.

The demonstratives are remarkably pervasive in speech. They serve a variety of purposes. As in many languages, they are used to distinguish previously mentioned referents, as in the examples above, and to introduce new referents, as in (11).

(11) New referent

Tánón’ *ki:* *Saksárie*,
 tanon’ *kiken* Saksárie
 and this NAME
o’nó:wa’ ne: *thaterennótha’*
o’now-a’ ne: t-ha-ate-renn-ot-ha’
 guitar that.one DV-M.SG.AGT-MID-song-stand-HAB
sok wahorihónnien’
sok wa-ho-rihw-onni-en’
 then FAC-M.SG>M.SG-matter-make-BEN-PFV
 ‘And he taught this guy Franklin to play the guitar.’

[Joe Awenhráthon Deer, speaker]

Demonstratives are also used pervasively as antitopics at the ends of clauses to confirm the identity of referents and often to mark conclusion of a thought.

(12) Antitopic

[‘Maybe you remember, they gave out horses to various people. Some were good animals, some were bad animals. You couldn’t hitch them up.’]

Iáh ki’ tekarihonnié:ni thi:.
iah ki’ te-ka-rihw-onni-en-i thiken
 not actually NEG-Z-matter-make-BEN-ST that
 ‘They weren’t trained, those [horses].’

[Joe Tiorhakwén:te’ Dove, speaker]

4 Placeholders

The Mohawk demonstratives also occur as fillers. Hayashi & Yoon (2006) characterize *fillers* as elements used “in contexts where speakers encounter trouble recalling a word or selecting the best word to use to designate some entity during the course of producing an utterance” (2006: 485). One type of filler is the *placeholder*, which they define as follows.

- It is a referential expression that is used as a substitute for a specific lexical item that has momentarily eluded the speaker (and which is often specified subsequently as a result of a word search).
- It occupies a syntactic slot that would have been occupied by the target word, and thus constitutes a part of the syntactic structure under constructions. (Hayashi & Yoon 2006: 490)

Hayashi & Yoon (2006) and others have observed that the sources of placeholders are often demonstratives.

Both of the Mohawk demonstratives *kí:ken* ‘this one, these’ and *thí:ken* ‘that one, those’ are used as placeholders in just the contexts described by Hayashi & Yoon (2006) when speakers are searching for how to designate an entity. Example (13) shows the placeholder *kí:ken*, ‘this one’, which fills the same syntactic slot as a referring expression which follows, ‘the boy going along with the stolen fruit’. (In transcriptions here shorter pauses are represented by sequences of two dots < . . >, and longer pauses by three dots < . . . >.)

- (13) Proximal demonstrative placeholder

<i>Rononhwaro'tsherónoint</i>	<i>kíken: . . um, . .</i>
ro-nonhwaro-’tsher-onti-on	kíken
M.SG.PAT-scalp.cover-NMLZ-lose-ST	this

‘He lost his hat **this** . . um, . . .

<i>raksà:'a</i>	<i>tsi niká:ien'</i>	<i>ne:</i>	<i>. . .</i>
ra-ksa'=a	tsi ni-ka-ien-'	ne	
M.SG-child=DIM	as PRT-N-lie-ST	the.aforementioned	

the [aforementioned] boy who . . . ,

rohianenskwenhátie'.

ro-ahi-a-nenskw-en-hatie'

M.SG.PAT-fruit-LK-steal-ST-PROG

the one who was going along with the stolen fruit.'

[Annette Kaia’titáhkhe’ Jacobs, speaker]

Example (14) shows the use of *thí:ken*, ‘that one’ as a placeholder, standing in for the nominal phrase *Variety Show*.

- (14) Distal demonstrative as placeholder

MB <i>Shaià:ta</i>	<i>ka'níhsera'</i>
s-ha=ia't-at	ka-'ni-hser-a'
REP-M.SG.AGT-body-be.one	N-be.father.to-NMLZ-NS

‘One father

<i>kwáh</i>	<i>ia'tahséntho'</i>	...
kwah	ia'-t-ha-ahsentho-'	
just	TRL-FAC-DV-M.SG.AGT-cry-PFV	

started to cry there,

thiken:-- . .

that one

<i>nahò:ten'</i>	<i>na'</i> —
na-h-o'ten'	na'—
PRT-N-be.a.kind.of-ST	

what is it—’

WG *Variety show*.

MB *Variety show*.

‘at *Variety show*’

[Mina Tewateronhiákhwa’ Beauvais, Warisóse Gabriel, speakers]

The demonstratives in these examples show all the characteristics of place-holders described by Hayashi & Yoon (2006):

The use of a placeholder demonstrative appears to be motivated by constraints in cognitive processes, such as difficulty in remembering or “accessing” an appropriate lexical item when it should be articulated during the course of utterance production. There, it is often accompanied by those features that are typically observed during word searches, such as **intra-turn pauses, sound stretches, repetitions, hesitation signals** (equivalents of *uh/um*), **self-addressed questions**, etc. (Hayashi & Yoon 2006: 500)

The demonstrative *kí:ken* in (13) was followed by a pause, then the hesitator *um*, then another pause. The demonstrative *thí:ken* in (14) was truncated, then

followed by a pause, then by *nahò:ten' na'* – ‘what is it wha-- ’, another sign of disfluency, as the speaker searched for a word.

The demonstratives in this function also show characteristic prosody: stressed penultimate syllable with rising pitch and added length, followed by final syllable with falling pitch. Basic word stress in Mohawk is penultimate. (Penultimate stress was established before the time of Proto-Northern-Iroquoian. Since then epenthetic vowels have been added in certain phonological and morphological contexts in Mohawk, which do not affect stress.) The primary acoustic correlates of stress are pitch and duration. Short stressed syllables show relatively high pitch, represented by an acute accent <á>. Open stressed syllables are lengthened. Lengthened stressed syllables show either rising pitch, represented by an acute accent and colon <á:>, or a rapid rise then fall in pitch to below the baseline, represented by a grave accent and colon <à:>. The basic pronunciation of the proximal demonstrative is thus *kí:ken*, as in Figure 1, and that of the distal demonstrative is *thí:ken*, as in Figure 2, shown here with their conventional spellings.

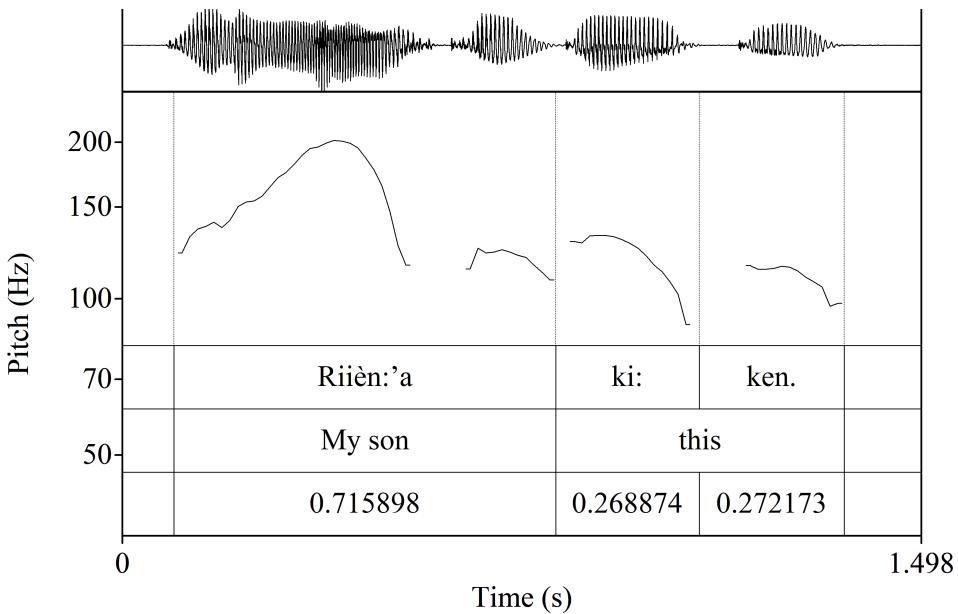


Figure 1: Basic proximal demonstrative *kí:ken*: ‘This is my son.’

In examples (13) and (14) above, in which the demonstratives were used as placeholders, there is little if any length or rise in pitch on the first syllable of the demonstratives, but, significantly, there is added length on the second

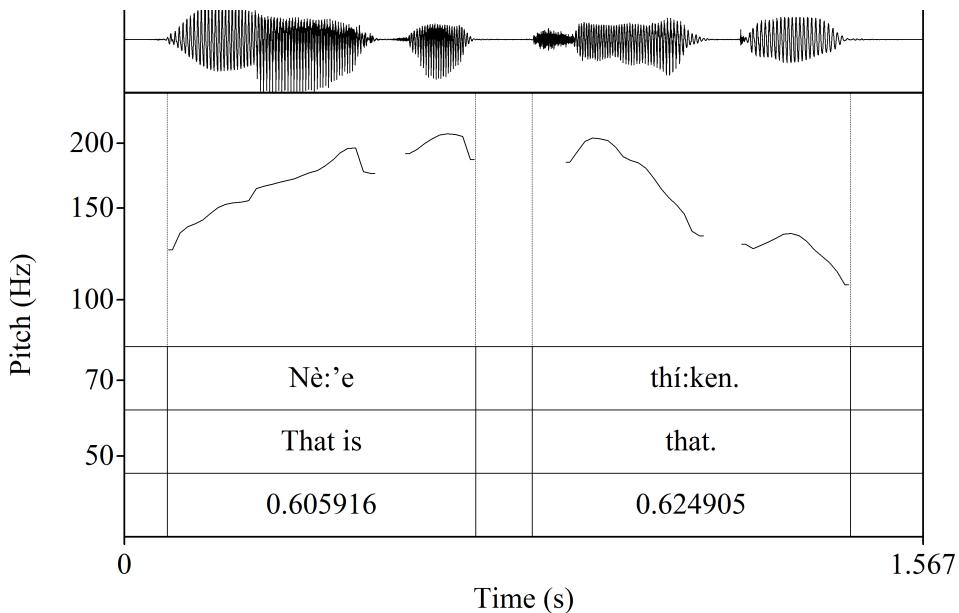


Figure 2: Basic distal demonstrative *thiken*: ‘It’s that one.’

syllable: *kiken*: ‘this one’, *thiken*: ‘that one’. The prosodic contour of (13) is shown in Figure 3.

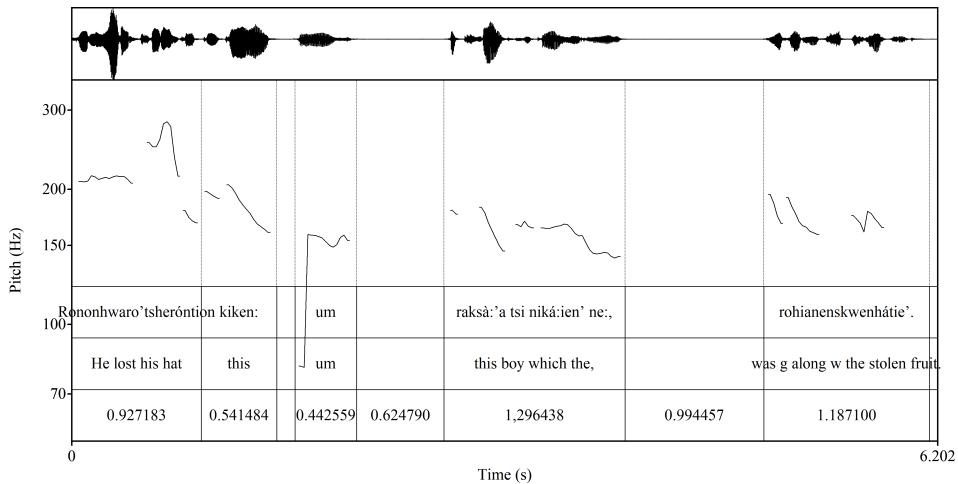


Figure 3: Placeholder *ki:ken* ‘this’, example (13)

A closer view of the prosody of the demonstrative is in Figure 4. The second syllable of the demonstrative ‘this’ is noticeably longer than the first.

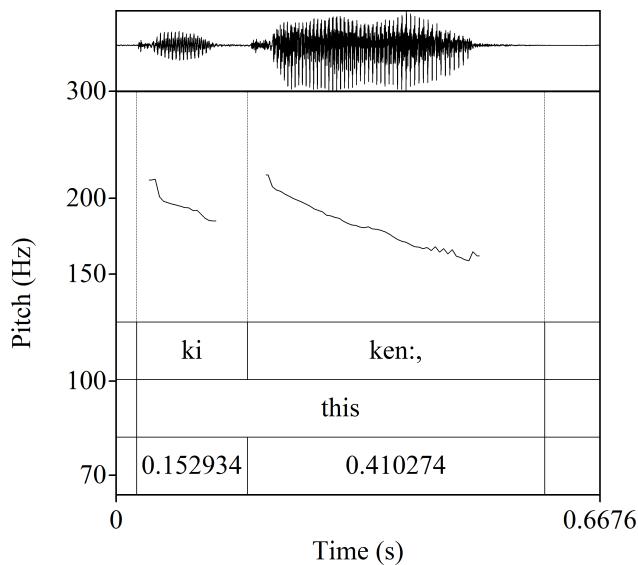


Figure 4: Closer view of demonstrative placeholder and following pause, example (13)

A similar pattern can be seen in the prosody of the placeholder construction with *thi:ken* ‘that’ from example (14) in Figure 5. Here, too, the second syllable of the demonstrative is noticeably longer than the first. A closer view shows the contour in more detail.

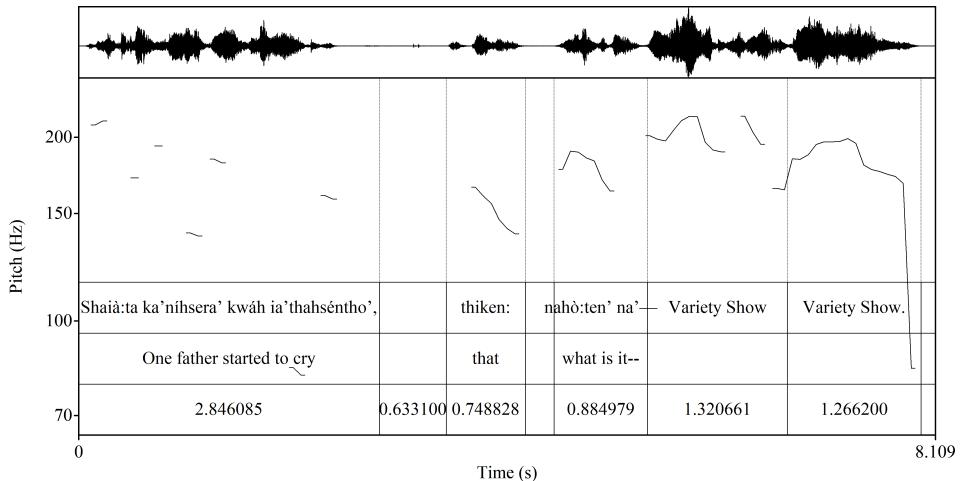


Figure 5: Placeholder *thi:ken* ‘that’, example (14)

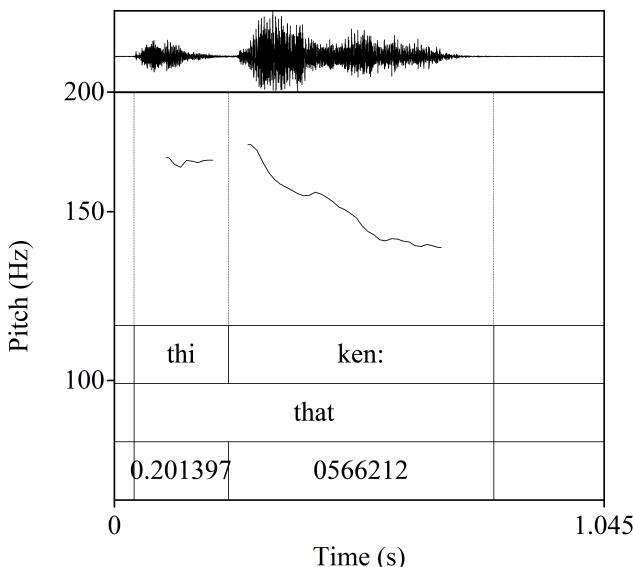


Figure 6: Closer view of demonstrative placeholder and following pause, example (14)

Like Hayashi & Yoon (2006), Podlesskaya notes that “a placeholder may fully or partially mirror the grammatical shaping of its target” (2010: 11). She adds (2010: 18) that languages may:

- exactly replicate the full grammatical marking of the delayed constituent
- not replicate the grammatical marking of the constituent
- allow partial replication

Neither Mohawk demonstratives nor nominals are inflected for number, gender, or case: this is a strictly head-marking language. The demonstratives in these constructions thus would not show the kind of mirroring of such features observable in some other languages in any case. The choice between the proximal and distal demonstratives in this Mohawk construction does, however, mirror the status of the referent of the following constituent in terms of distance in place, time, discourse mention, etc. The proximal demonstrative in (13), ‘He lost his hat **this** . . . [aforementioned boy going along with the stolen fruit]’ referred to the central protagonist at that point in the narrative, a core argument of most sentences leading up to this one. The distal demonstrative in (14), ‘One father started to cry

at that . . . [Variety Show]', referred to a location that had not been mentioned before and was more remote in place and time. Beyond this, the only basis for judging whether an item is a placeholder is Hayashi & Yoon's (2006) characterization, "a referential expression that is used as a substitute for a specific lexical item".

5 Interjective hesitators

A second type of filler described by Hayashi & Yoon (2006) is the *interjective hesitator*. It is distinct from the placeholder in that the demonstrative does not point to a particular kind of referent. While placeholders are produced as a syntactic constituent of a sentence, hesitators are not (2006: 489). Mohawk demonstratives occasionally occur in such contexts of hesitation.

- (15) Demonstrative as interjective hesitator
Tánon' sewatié:rens ò:ni' né:' kwáh **ki:** *a*;
and sometimes also it.is just this ah
enthotinà:khwen'.
en-t-hoti-na'khwen'
FUT-CSL-M.PL.PAT-get.angry-PFV
'And sometimes they'll just get mad.'

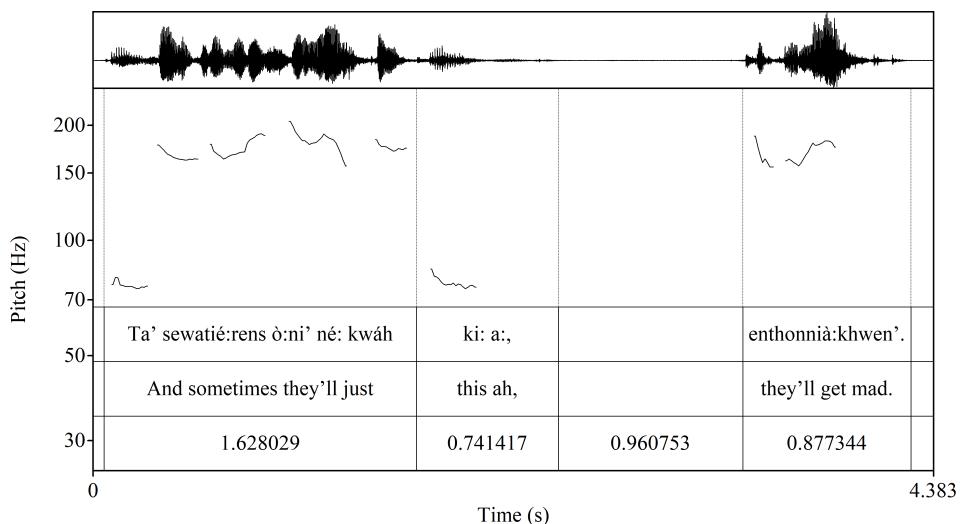
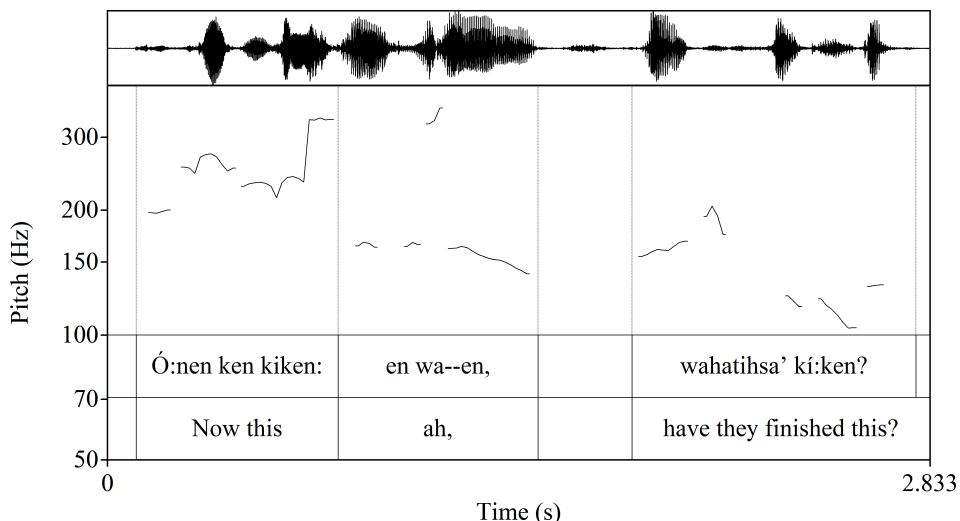
[Warisóse' Gabriel, speaker]

The demonstrative occurred before the filler *a*: then a pause, before the speaker continued with the verb 'they'll get mad', not a coreferential nominal.

Another example is in (16). The demonstrative occurred before the filler *en* then a false start *wa--* and another filler *en*, and a breath, before the speaker continued with the verb 'they finished this'.

- (16) Demonstrative as interjective hesitator
Ó:nen *ken* **kiken:** *en* *wa--en*,
now Q this ah wa--ah
wahatihsa' *kí:ken?*
wa-hati-ihsa-' kiken
FAC-M.PL.AGT-finish-PFV this
'Now this ah ah, have they finished this?'

[Doris White, speaker]

Figure 7: Interjective hesitator *kí:ken* ‘this’, example (15)Figure 8: Interjective hesitator *kí:ken* ‘this’, example (16)

The prosody can be seen in Figure 8 (p. 355).

In neither (15) nor (16) does the demonstrative anticipate the syntactic category of what follows.

In these interjective hesitator constructions, the demonstrative shows the same prosody as in the placeholder constructions, with added length on the final syllable and sometimes a continuing rise in pitch. This pattern is characteristic of continuing intonation unit patterns across the language when the penultimate syllable of the final word is open. In Figure 8 above, for example, the initial particle *ó:nen* ‘now’ was pronounced with added length and rising pitch on the final syllable (*ó:nen* -> *onén:*) as was the final syllable of the demonstrative (*kí:ken* -> *ki:kén:*).

Similar prosodic differences between basic demonstratives and fillers have been observed by Vallejos Yopán (2023) with a filler *este* in Peruvian Amazonian Spanish:

Filler-*este*, which originally evolved from demonstrative-*este*, serves to deal with word-formulation delays during spontaneous speech production. Analyses of conversations reveal that *este* primarily functions as a filler: 70% of the tokens of *este* are either fillers serving as placeholders, which replace lexical items in specific syntactic slots, or fillers serving as hesitators, which are non-referential and distributionally free. Further, phonetic analyses show that demonstrative-*este* and filler-*este* exhibit different phonetic shapes. Demonstrative-*este* patterns with disyllabic words with penultimate stress – the first vowel is longer than the second vowel. Filler-*este* shows the opposite configuration – the second vowel is significantly longer than the first vowel. (Vallejos Yopán 2023: 651)

Vallejos Yopán and others have pointed out that the development of fillers from demonstratives is not entirely surprising. “One scenario for the emergence of the filler use [of demonstratives] could be the pronominal use with cataphoric reference, if the identity of *este* is specified in the following discourse.” (2023: 671) She further proposes the following:

The repetition of *este* ‘this one’ in this context seems natural. This path of development entails extending the use of *este* from pointing to entities in space, entities presented in prior discourse, or those to be specified cataphorically, to indicate issues with lexical retrieval. Next, the form employed as a placeholder becomes invariant, it no longer holds gender or number agreement with the referent. Truncated placeholders over time become signs of

hesitation and loose referentiality, thus becoming interjective hesitators.
(2023: 671)

The use of demonstratives as hesitators in Mohawk is actually not as common as their use in another function, however.

6 Beyond fillers

This structure can be seen in (17). As in the filler constructions, the demonstrative here showed extra lengthening of the final syllable and a non-terminal pitch contour, then was followed by a pause before the next intonation unit, which began with a partial pitch reset.

- (17) Discourse structure

Thó ki' wahonwéntskaron'se' kikén; . . .
there in fact wa-honwa-itskar-on-'s-e' kiken
there in fact FAC-FI>M.SG-rug-make-BEN-PFV this
'There she made him a pallet, ...

tsi tsiiotékha'.
tsi io-atek-ha'
place N.PAT-burn-HAB
by the fire.'

[Sadie Sesír Smoke Peters, speaker]

The prosody can be seen in Figure 9 (p. 358).

The same pattern can be seen in (18). The first intonation unit ended in a demonstrative with added length on the final syllable and non-terminal pitch contour, followed by a pause and partial pitch reset on the following intonation unit.

- (18) Discourse structure

Wa'tiakwatska'hòn:ne' kikén; . . .
wa'-t-iakwa-atska'hon-hne' kiken
FAC-DV-1EXCL.PL.AGT-dine-PURP this

Joyce Mitchell.

'We all were going to eat with Joyce Mitchell.'

[Watchenní:ne' Sawyer, speaker]

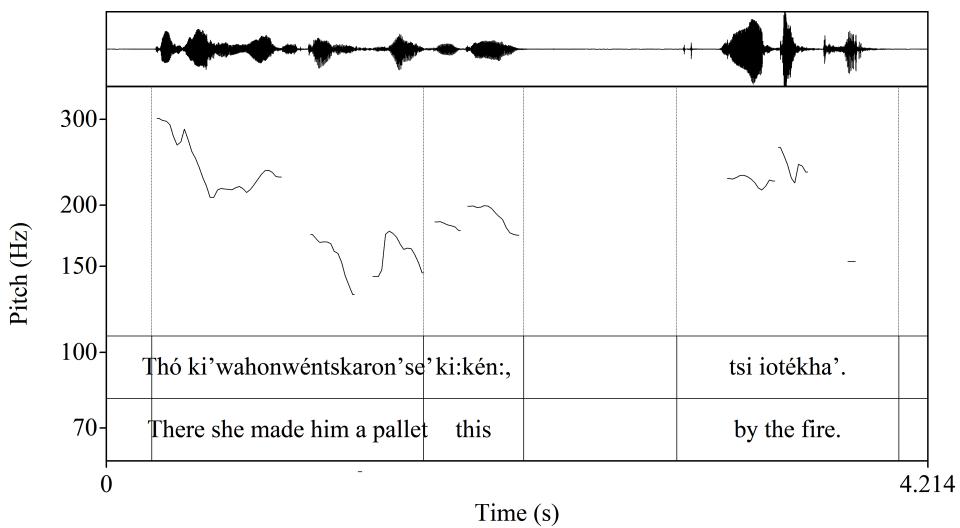


Figure 9: Discourse construction with *kí:ken* 'this', example (17)

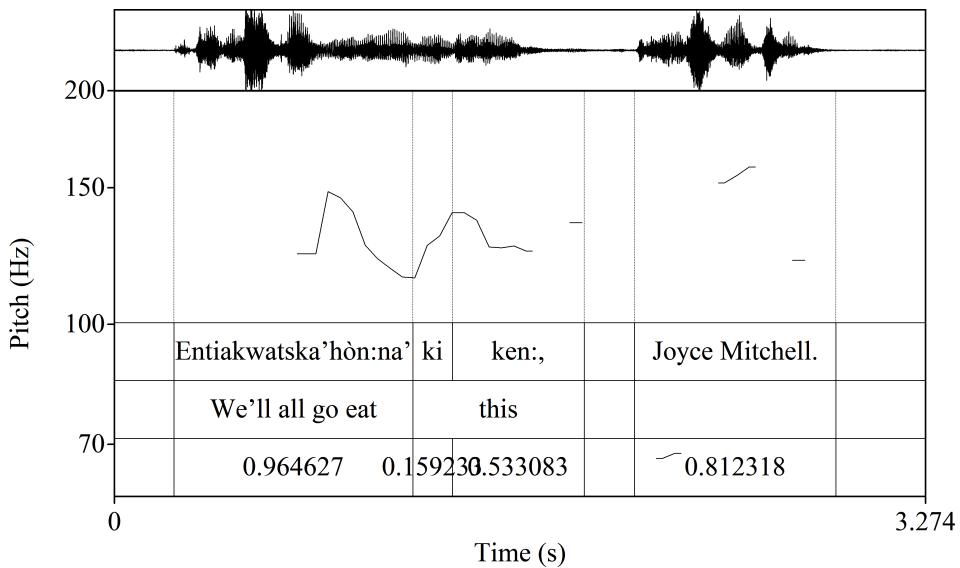


Figure 10: Discourse construction with *kí:ken* 'this', example (18)

The prosody can be seen in Figure 10 (p. 358).

Like Hayashi & Yoon's (2006) interjective hesitators, the demonstratives here are not necessarily coreferential with what follows. Speakers confirm that the demonstrative *kí:ken* 'this' in (19) did not refer to the week.

- (19) Mohawk discourse pattern

<i>Ó:nare'</i> , . . .	
onen	á:re'
now	again
<i>saionkenikè:tohte'</i>	<i>kikén:</i>
sa-ionkeni-ke'toht-e'	<i>kiken</i>
REP.FAC-1PL.PAT-appear-PFV	this
<i>ó:nen tsiahià:kshera</i>	[. . .]
onen ts-i-ahia'k-sher-at	
now REP-N-week-NMLZ-be.one	
'Now we're back again after one week.'	

[Joe Awenhráthon Deer, speaker]

But these uses differ in a significant way from those of fillers: they are not associated with disfluency. The speakers were not searching for words or hesitating about how to formulate the next idea. They are part of a larger phenomenon.

A number of authors have observed that speech is not produced sentence by sentence, but in spurts, termed by Halliday (1967) "information units", by Crystal (1975) "tone units", by Grimes (1975) "information blocks", by Kroll (1977) "idea units", and by Chafe (1980, 1994, 2018) "intonation units". These units are characterized by patterns of pitch, intensity, rhythm, voice quality, and pausing. Chafe pointed out that "these spurts of language are linguistic expressions of focuses of consciousness" (1980: 15) and that "each intonation unit verbalizes the information active in the speaker's mind at its onset". (1994: 63)

Hayashi & Yoon (2006) proposed that the use of demonstratives as interjective hesitators stems from cognitive factors much like those described by Chafe:

We consider the possibility that demonstratives in this usage have undergone the process of "pragmaticization". We will suggest that the demonstratives used as interjective hesitators have been pragmaticized for the discourse function of monitoring the flow of the production of an utterance to such an extent that they have diverged from ordinary demonstratives in terms of syntactic distribution, referentiality, and correspondence between morphological forms and semantic meanings. (Hayashi & Yoon 2006: 488)

The Mohawk discourse structure has developed as a conventionalized way of packaging information for both the speaker and the hearer. It can facilitate interaction. The demonstrative at the end of an intonation unit, pronounced with a non-terminal prosodic contour, signals to listeners that more is to come. It can thus serve as a floor-holding device. It can also facilitate processing by listeners, allowing them to take in one new idea at a time.

Keevallik (2010) describes strikingly similar effects in Estonian. The Estonian demonstrative *see* serves as a placeholder during word searches, but Keevallik observes that it also does more. She points to its role in facilitating smoother interaction:

Why a delay needs to be announced has generally been explained from the perspective of the speaker and her cognitive limitations. The reasons have included a planning or formulation problem, memory search, doubt, uncertainty or hesitation (see the summary in Clark & Fox Tree 2002: 90). It has been assumed that the speaker is in general unable to proceed with talk at the current moment. In contrast, this study will explore the possibility that a filler may be implemented as a conscious strategy for achieving certain interjective ends. (Keevallik 2010: 139)

She also points to the facilitation of processing for the listener:

There may be interpersonal advantages in the delay of key items in turns, such as easing the processing for the recipient, announcing structural boundaries in conversation, and displaying orientation to the sensitiveness of the action. (2010: 139)

As she notes, both Hayashi & Yoon (2006: 526–527) and Schegloff (2010) observe that fillers can mark larger structural boundaries in conversation, often prefacing an introduction to a new topic. She finds similar patterns in her Estonian data, drawn primarily from a corpus of 324 telephone calls, augmented with some face-to-face conversation:

Topic initiations in general tend to be accompanied by delays, and reasons for the call will as a rule initiate new topics in conversations. Accordingly, the Estonian *see* is regularly used when presenting reasons for the call. (2010: 170)

Kosmala & Crible (2022: 216) found similar patterns with fillers *euh* and *eum* in spoken French: “Although they [filled pauses] are commonly associated with

hesitation, disfluency, and production difficulty, it has also been argued that they can serve more fluent communicative functions in discourse (e.g., turn-taking, stance-marking).” They cite previous work indicating that in addition to their association with hesitation, filled pauses can also mark discourse structure (Swerts 1998) and manage turn-taking (Kjellmer 2003, Beňuš 2009).

Just such effects can be seen with the Mohawk demonstrative structure. The speaker in (20) was opening a brand new topic of discussion, his garden. The first intonation unit ended with the demonstrative with characteristic prosody, with higher pitch and length on the stressed penultimate syllable *ki:* then falling pitch on the ultimate syllable. The demonstrative did not refer cataphorically to the following constituent, the summer; if it referred to anything, it was the whole new topic of conversation.

- (20) Mohawk discourse pattern

Né: kiken:,

ne:’=e kiken

that=is this

nòn:wa’ akenhná:te’ né:ne,
nonhwa’ akenh-at-e’ ne’ne
now summer-stand-ST which

tiotieriénhton

t-io-at-iernenht-on

CSL-N.PAT-be.first-ST

iáh teiotón:’on thé:nen aontío’ten’.
iah te-io-aton’-on othenen aa-wak-io’t-en’
not NEG-N.PAT-be.poss-ST anything IRR-1SG.PAT-work-PFV
‘This summer was the first time I wasn’t able to work.’

[Joe Awenhráthon Deer, speaker]

This Mohawk pattern, in which one intonation unit ends in a demonstrative pronounced with added length on the final syllable, a non-terminal prosodic contour, and often a pause, is now pervasive in speech, but the frequency varies widely across speakers, topics, and contexts. In one conversation involving five speakers, 50% of the demonstratives *ki:ken* ‘this/these’ and *thí:ken* ‘that/those’ occurred in these structures. In another conversation involving two of the same speakers plus another, the figure was 39%. In a third, involving two other speakers, the figure was 30%. In some others, it was considerably less. An important point is that there is not a sharp distinction between the filler pattern and the

discourse pattern; technically the only difference would be whether it is a result of disfluency.

Similar patterns can be seen in related Northern Iroquoian languages, though the demonstratives recruited are not necessarily cognate. The Tuscarora counterparts *kyè:ní:kę*: ‘this/these’ and *hè:ní:kę*: ‘that/those’ are used in the same ways, as are the Cayuga counterparts *nę:gyęh* ‘this/these’ and *to:gyęh* ‘that/those’.

7 Toward syntax?

It is well known that cross-linguistically, complementizers and relativizers are often descended from demonstratives (Brugmann 1904, Bühler 1934, Hopper & Traugott 2003, Harris & Campbell 1995, Diessel 1999, 2006, Heine & Kuteva 2007, Diessel & Breunesse 2020, among many). These are typically analyzed synchronically as part of the dependent clauses: *It is regrettable [that our language is disappearing], I remembered [that we are striving to speak], It is different for the children [that grew up there]*. The sources of these constructions have been hypothesized to be sequences of sentences, one of which contained a demonstrative. A continuing topic of discussion has been whether the demonstrative was originally part of what became the matrix clause or the dependent clause, and whether it was cataphoric or anaphoric. Diessel & Breunesse (2020: 308–310, 319–320) trace some of the history of discussion.

In many cases, it is not possible to identify the precise source construction on the basis of written records alone. Mohawk provides a glimpse of a point along one possible pathway of development. Complement and relativization constructions have not crystallized in Mohawk to the same extent as those in some other languages. In fact their counterparts have the same structure as that of the place-holder, hesitator, and discourse-structuring pattern seen in the previous sections.

7.1 Toward complement constructions?

The Mohawk pattern seen in previous sections is often used in situations in which complement constructions would be used in other languages. An idea is presented in one intonation unit ending in a demonstrative with non-final prosody, which indicates that further elaboration is to follow. Some occurrences of this structure are comparable to subject complement constructions in other languages, like that in (21). In the Mohawk, the demonstrative is grouped prosodically with what would be the matrix clause, however, while in the English, it is analyzed as part of the complement.

- (21) Toward a subject complement?

Sénha' ki' ietiohná:ten **kikén:** . . .

senha' ki' ie-t-io-nhat-en kiken

more in.fact TRL-DV-N.PAT-regret-ST this

more in fact it is regrettable this

onkwehonwehnéha' onkwawén:na' iohtentionhátie'.

onkweh=onweh=neha' onkwa-wenn-a' io-ahtenti-on-hatie'

person=trad=style 1PL.AL.POSS-lg-NS N.PAT-move-ST-PROG

Native our language it is moving along

'It is all the more regrettable [that our Indian language is getting lost].'

[Watshenní:ne' Sawyer, speaker]

The prosody can be seen in Figure 11, in which the first intonation unit ended in the proximal demonstrative *kiken*: 'this one' with added length and rising pitch on the final syllable. The next clause began after a pause.

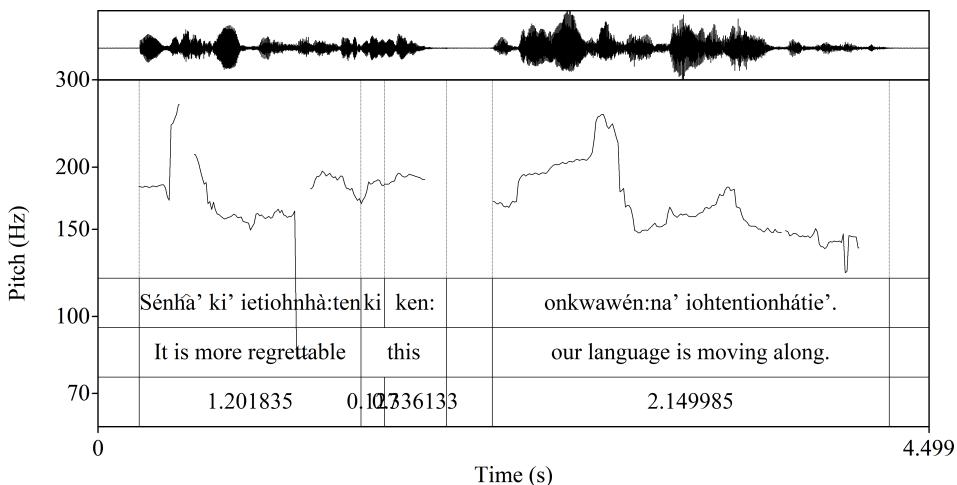


Figure 11: Toward a complement construction? Example (21)

The sentence in (22) shows a similar pattern. The first intonation unit ended in the distal demonstrative *thí:ken* 'that one' with final lengthening, indicating that more was to follow. The next began after a pause and provided further elaboration.

- (22) Toward a subject complement?

Ó:nen ni' énska ia'ká:ienhte' *thiken*; . . .
 onen ohni' enskat i-a'-ka-ienht-e' thiken
 now also one TRL-FAC-N.AGT-hit-PFV that
 now also one there it hit that

 thó ki' á:re' ní:ioht
 tho ki' are' ni-io-h_t
 there in.fact again PRT-N.PAT-be.so
 there in fact again so it is

 ó:ia' ia'tekohtáhrhohs.
 o-i-a' ia'-te-k-ohtarho-hs
 N-other-NS TRL-DV-1SG.AGT-clean-HAB
 other I was cleaning there
 'Then once again it happened
 [that I was cleaning in there again].'

[Kahentoréhtha' Marie Cross, speaker]

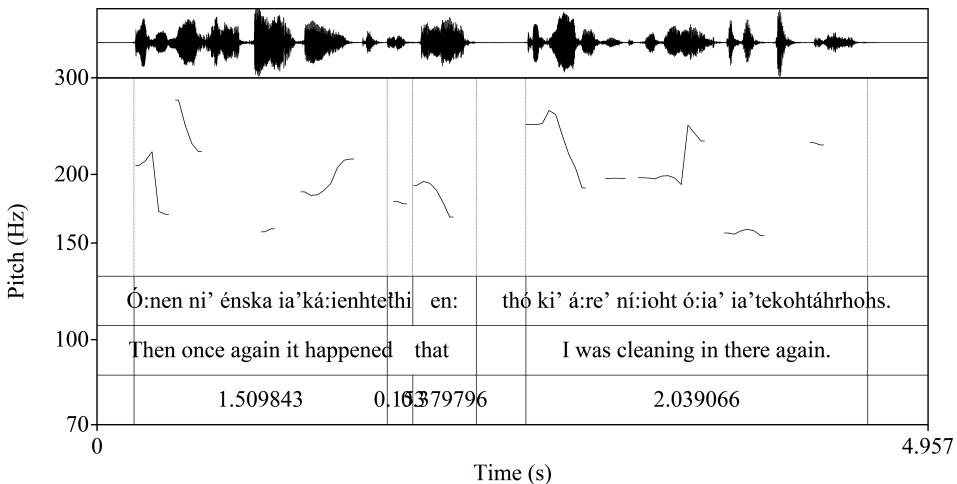


Figure 12: Toward a complement construction? Example (22)

The same pattern is used in contexts where object complement constructions are used in some other languages. One intonation unit ends in a demonstrative showing added length and sometimes rising pitch on the final syllable, indicating that further elaboration is to follow. The next begins after a pause with a

partial pitch reset. An example is in (23). (Neuter arguments are not marked explicitly in the pronominal prefixes on verbs unless there is no other argument, but transitivity is implied by the verb.)

- (23) Toward an object complement?

<i>Né:</i>	<i>ki'k</i>	<i>wa'kehià:ra'ne'</i>	<i>kiken;</i> . . .
ne'e ki'=ok	wa'-k-ehiahr-a'n-e'	<i>kiken</i>	
it.is in.fact=just	FAC-1SG.AGT-remember-INCH-PFV	this	
<i>tsi niió:re'</i>	<i>tsi kiótkon ionkwahskéhnen,</i>		
tsi ni-io-r-e'	tsi tiotkon ionkwa-ahskehnh-en		
as PRT-N-extend-ST how always 1PL.PAT-strive-ST			
<i>nonkwawén:na',</i>			
ne=onkwa-wenn-a'			
ART=1PL.AL.POSS-language-NS			
<i>onkwehonwehnéha',</i>			
onkweh=onweh=neha'			
person=traditional=style			
<i>aonsetewatá:ti'.</i>			
a-ons-etewa-atati-			
IRR-REP-1INCL.PL.AGT-speak-PFV			
'I just remembered [how we are always striving to still speak our Native language].'			

[Sadie Sesir Smoke Peters, speaker]

The prosody can be seen in Figure 13.

The grouping of a demonstrative prosodically with the intonation unit before a clause which could function as a complement is not unique to Mohawk. In his discussion of asymmetries in the prosodic phrasing of function words, Himmelmann points out that “truncations after preposed function words are probably possible and attested in many, if not all, languages that have such function words” (2014: 936) and, furthermore, that “preposed function words are often chunked with the preceding word rather than with their morphosyntactic host, which follows the discontinuity” (2014: 937). Such chunking makes sense when prosody is understood in terms of foci of consciousness. Himmelmann cites the example in (24) from German, which he characterizes as evincing a type of disfluency. The complementizer *dass* ‘that’ appeared in the same intonation unit as the matrix verb rather than the complement.

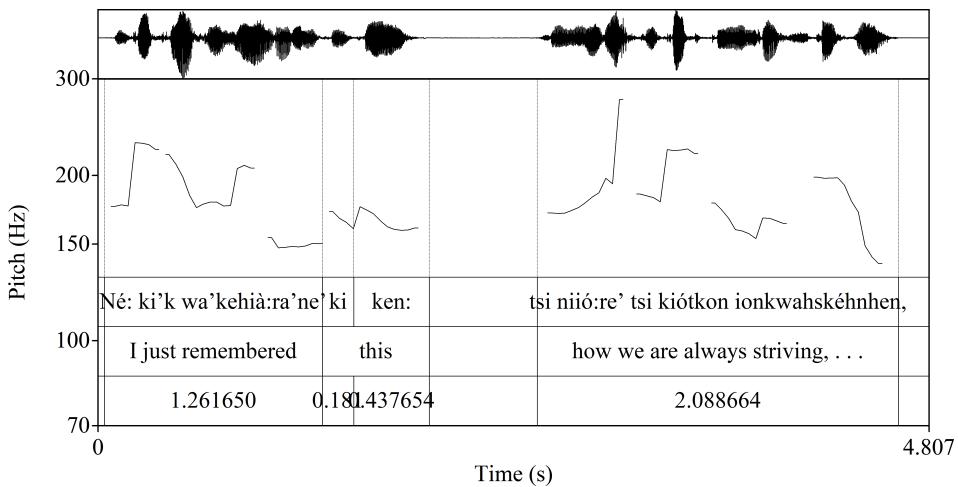


Figure 13: Toward a complement construction? Example (23)

- (24) German prosody: Himmelmann 2014: 937

un: dann hab' ich plötzlich von weitem,
and then have I suddenly from afar

(0.5) *gesehen dass,*
seen that

(0.8) *en Teil von der—*
a part of the

(0.8) *öh öh (0.4) Strecke,*
uh uh road

öh mit Schnee,
uh with snow

öh (0.8) ähm (1.1)
uh um

also (0.4) mit Schnee bedeckt war.
well with snow covered was

7.2 Toward relative clause constructions?

Somewhat less common in Mohawk is the use of the pattern in contexts where relative clauses might occur in other languages. One intonation unit ends in a demonstrative with added length and a non-terminal pitch contour, indicating

that more is to follow. The next begins after a pause and provides further elaboration.

- (25) Headless relative clause?

Shakorihonnién:ni kwi' thi:kén:, . . .
 shako-rihw-onné-enni ki'=wahi' thiken
 M.SG>FI-matter-make-BEN in факт=TAG that
 he makes words for us in fact that

shonkwahnhà:'on.
 shonkwa-nha'-on
 M.SG>1PL-hire-ST
 he hired us
 'The teacher is the one [that told us to].'

[Kahentoréhtha' Marie Cross, speaker]

The prosody can be seen in Figure 14.

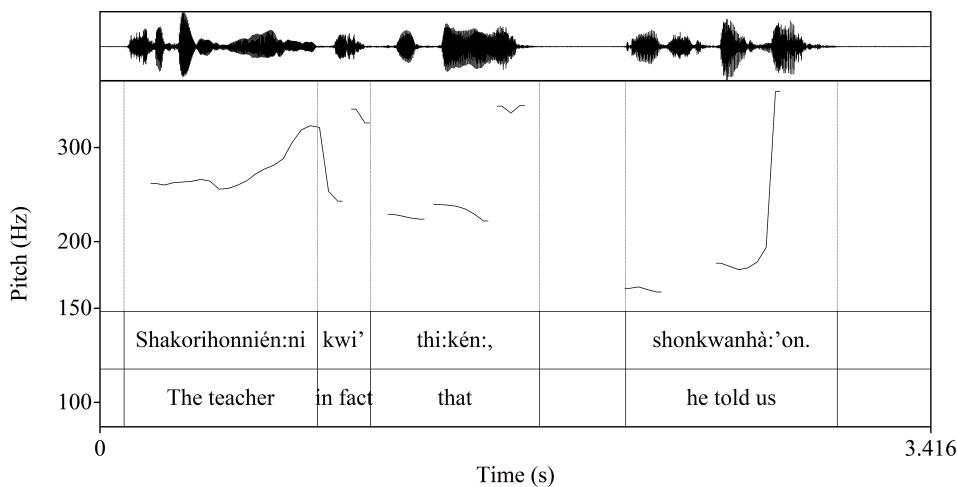


Figure 14: Toward a headless relative clause? example (24)

Though less frequent, examples occur that might approach the form of headed relative clauses.

- (26) Relative clause?

Né: aori:wa' ki:kén:,
 ne: ao-rihw-a' kiken
 it.is N.AL.POSS-reason-NS this
 it is its reason this

kkaraien té:ri akwé:kon.
k-kar-a-ienteri akwekon
1SG.AGT-story-LK-know.ST all
I know the story all
‘That is the reason [that I know the whole story].’

[Lazarus Jacob, speaker]

- (27) Relative clause?

Wakerihwaiè:wahse' se's ka' nón: ki:kén;
wake-rihw-a-iehwahs-e' se's ka' nonwe kiken
1SG.PAT-reason-LK-be.unable.to.find-ST then what place this
thonathéhtaien' né: ki: né;
t-hon-at-heht-a-iен' ne'e kiken ne'e
CSL-M.PL.PAT-MID-field-LK-have-ST it.is this it.is
rake'niha tsi nonkwá:ti.
rake-'ni=ha tsi nonkwati
M.SG>1SG-be.father.to=DIM place side
‘I couldn't figure out the place
[where their garden was on my father's side].’

[Joe Tiorhakwén:te' Dove, speaker]

8 Conclusion

Many of the features observable in the placeholder and hesitator structures in Mohawk are also characteristic of a more general discourse pattern in the language. Hayashi & Yoon (2006) cited similar features of their interjective hesitators, which they described as “pragmaticized for the discourse function of monitoring the flow of the production of an utterance to such an extent that they have diverged from ordinary demonstratives in terms of syntactic distribution, and referentiality” (2006: 488). The Mohawk discourse construction serves as a general strategy for managing the flow of information. It consists of an intonation unit ending in a demonstrative with added length on the final syllable and a non-terminal pitch contour, often followed by a pause, then followed by another intonation unit beginning with a partial pitch reset which provides some elaboration. Each intonation unit expresses a new idea, a single focus of consciousness in the sense of Chafe (1980, 1994). In some cases, in the placeholder and filler uses, the pattern appears with observable disfluency, but in many more, there is no indication of disfluency at all.

The demonstrative in this discourse construction may be understood to refer cataphorically to what follows or not, reflecting the divergence cited by Hayashi & Yoon (2006) for hesitators “from ordinary demonstratives in terms of syntactic distribution and referentiality”. Reference is cataphoric in the placeholder uses and some general discourse uses, including those comparable to complement and relative clauses in other languages, but not (simply by definition) in hesitator uses and in many other discourse uses.

It is perhaps not surprising that such constructions should develop in languages in general. They facilitate processing for both speaker and listener. As is well known, speakers tend to present one significant new idea at a time prosodically, in successive intonation units. Such phrasing can not only allow speakers time to formulate the next idea, but also to organize the information for their listeners. It can signal larger discourse structure, occurring at major discourse breaks. It can also facilitate interaction, signaling to listeners that the speaker wishes to hold the floor. As more audio documentation of unscripted, interactive speech in more languages becomes available, it is likely that similar patterns will be observed more widely.

Abbreviations

AGT	grammatical agent	M	<u>masculine</u>
AL	alienable	MID	middle
ART	article	N	neuter gender
BEN	benefactive applicative	NMLZ	nominalizer
CAUS	causative	NS	noun suffix
COIN	coincident	PAT	grammatical patient
CSL	cislocative	PFV	perfective aspect
DIM	diminutive	PL	plural
DV	duplicative	POSS	possessive
EXCL	exclusive	PROG	progressive
FAC	factual	PROH	prohibitive
FI	feminine/indefinite gender	PRT	partitive
FUT	future tense	PURP	purposive
HAB	habitual aspect	Q	question particle
INCH	inchoative	REP	repetitive
INCL	inclusive	SG	singular
IRR	irrealis	ST	stative aspect
LK	linker	TRL	translocative

References

- Beňuš, Štefan. 2009. Variability and stability in collaborative dialogues: Turn-taking and filled pauses. In Maria Uther, Roger Moore & Stephen Cox (eds.), *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, 796–799. Grenoble: International Speech Communication Association (ISCA). DOI: 10.21437/Interspeech.2009-180.
- Brugmann, Karl. 1904. *Die Demostrativpronomina der indogermanischen Sprachen: Eine bedeutungsgeschichtliche Untersuchung*. Leipzig: B. G. Teubner.
- Böhler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.
- Chafe, Wallace L. 1980. The deployment of consciousness. In Wallace L. Chafe (ed.), *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*, 9–50. Norwood: Ablex.
- Chafe, Wallace L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chafe, Wallace L. 2018. *Thought-based linguistics: How languages turn thoughts into sounds*. Cambridge: Cambridge University Press.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Crystal, David. 1975. *The English tone of voice: Essays in intonation, prosody and paralanguage*. London: Edward Arnold.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization* (Typological Studies in Language 42). Amsterdam: John Benjamins.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4). 463–489. DOI: 10.1515/COG.2006.015.
- Diessel, Holger & Merlijn Breunesse. 2020. A typology of demonstrative clause linkers. In Åshild Næss, Anna Margetts & Yvonne Treis (eds.), *Demonstratives in discourse*, 305–341. Berlin: Language Science Press. DOI: 10.5281/zenodo.4055834.
- Grimes, Joseph E. 1975. *The thread of discourse*. The Hague: Mouton.
- Halliday, M. A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244. DOI: 10.1017/S002226700016613.
- Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of

- word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar: A reconstruction* (Oxford Studies in the Evolution of Language). Oxford: Oxford University Press.
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960. DOI: 10.1353/lan.2014.0105.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. 2nd edn. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Kjellmer, Göran. 2003. Hesitation. In defence of *er* and *erm*. *English Studies* 84(2). 170–198. DOI: 10.1076/enst.84.2.170.14903.
- Kosmala, Loulou & Ludivine Crible. 2022. The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1). 216–239. DOI: 10.1177/00238309211010862.
- Kroll, Barbara. 1977. Combining ideas in written and spoken English: A look at subordination and coordination. In Elinor Keenan & Tina Bennett (eds.), *Discourse across time and space* (Southern California Occasional Papers in Linguistics 5), 69–108. Los Angeles: University of Southern California.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Schegloff, Emanuel A. 2010. Some other “uh(m)”s. *Discourse Processes* 47(2). 130–174. DOI: 10.1080/01638530903223380.
- Swerts, Marc. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30(4). 485–496. DOI: 10.1016/S0378-2166(98)00014-9.
- Vallejos Yopán, Rosa. 2023. From demonstrative to filler: *Este* in Amazonian Spanish and beyond. *Linguistics* 61(3). 651–678.

Chapter 11

Mashti: A multipurpose filler in Northern Pastaza Kichwa

Alexander Rice
University of Alberta

This chapter presents a multimodal analysis of the filler word *mashti* and its functions in Northern Pastaza Kichwa, a Quechuan language spoken in the Amazonian lowlands of Ecuador and Peru. In discourse, *mashti* occurs less frequently than the language's other fillers, but it has some unique properties that set it apart from the others. Unlike the language's discourse interjection and demonstrative-derived fillers, *mashti* derives from a lexicalized "whatchamacallit" type construction consisting of an interrogative *ima* 'what' and the noun *shuti* 'name'. In discourse, *mashti* can serve three functions: as a non-referential and distributionally unrestricted hesitative filler, as a nominal and verbal placeholder filler (which can fully mirror the morphology of the target word), and as a relexicalized pro-verb that reiterates a previously uttered action or event. These functions are corroborated not only by their morphosyntactic properties, but also by their relative co-occurrence with bodily correlates of disfluency and hesitation, such as prosodic lengthening, and the deactivation/stalling of manual gestures. Thus, this chapter contributes to and expands the typological research on hesitative and placeholder words by describing a multifunctional filler through a novel approach that takes into account the multimodal and prosodic signals associated with its different discourse functions.

keywords: Quechuan languages, hesitator, placeholder, multimodality, gestures, prosody

1 Introduction

Typological descriptions of minority languages have historically relied on "clean" language data and therefore prioritized samples of monologic language com-



Alexander Rice. 2025. *Mashti: A multipurpose filler in Northern Pastaza Kichwa*. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 373–439. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697595

posed of tidy and fluently produced utterances. More recently, studies in language documentation and cognitive linguistics have come to recognize the primacy of natural and contextualized language data and of language function. Research in this vein thus places a premium on language data coming from people immersed in real-life situations regardless of how “messy” or “non-standard” such data may be given their importance to the understanding of language as an emergent phenomenon rooted in social interaction.

One of the hallmarks of naturalistic speech is disfluency, or problems that arise in the planning and execution of speech (Lickley 2015: 452). To manage disfluencies, speakers make use of a range of discourse strategies which include hesitations, pauses, false starts, repairs, and prosodic lengthening. One such strategy is the use of **FILLER** words, which speakers use to buy time while experiencing difficulty remembering a certain word or content to express (Fox 2010).

In this chapter, I describe the uses of the filler word *mashti* in Northern Pastaza Kichwa, an indigenous language of South America spoken in the Ecuadorian-Peruvian Amazon. I show that speakers of Northern Pastaza Kichwa use the filler *mashti* as a **HESITATOR** (HES) or as a lexical **PLACEHOLDER** (PH). Furthermore, *mashti* can be used as **RESUMPTIVE PRO-VERB** (PROV). These functions of *mashti* are largely determined by the way it relates to the surrounding discourse. Crucially, this analysis is also multimodal and takes into account phenomena in the vocal-aural and visuo-spatial modalities. Specifically, I show that prosodic lengthening and gestural disfluency may be correlated with the use of *mashti* as a filler. Prosodic lengthening, in particular, is crucial in distinguishing between the hesitator and other uses of *mashti* when the degree of morphosyntactic integration is not obvious. Finally, the lexical origins and synchronic distribution of *mashti* functions suggest a multipronged path of development.

In the remaining sections of this chapter, I present the language and the nature of the data used for this study (§2.1)-(§2.2) and offer primers on fillers (§2.3) and multimodal analysis (§2.4). Subsequently, I describe the functions of *mashti* with contextualized examples (§3), as well as the co-occurring multimodal disfluency phenomena (§4). In §5, I discuss the origin and possible pathways of the development of *mashti* and conclude with a brief overview of the findings of this chapter in §6.

2 Preliminaries

The objective of this chapter is to explore how the word *mashti* in Northern Pastaza Kichwa is used as a filler and pro-verb in discourse. In service of this objective, this section sets up the background information and essential concepts.

In §2.1 I provide a brief introduction to Northern Pastaza Kichwa and the context its speakers inhabit. In §2.2 I sketch the outlines of the corpus used for this chapter as well as the nature of the data. In §2.3 I outline the concepts of filler words and the criteria used to distinguish between hesitative fillers and placeholder fillers. Finally, in §2.4, I present the basic tenets of multimodal analysis and illustrate how I represent multimodal elements in the linguistic examples used in this chapter.

2.1 Northern Pastaza Kichwa and its speakers

Northern Pastaza Kichwa is a member of the fairly large and diverse Quechuan language family, which is native to the Andean mountain range of South America. Northern Pastaza Kichwa belongs to the northern branch of the Quechuan languages, Quechua II-B (Torero 1964) spoken in the northern Peruvian Amazon, Ecuador and southern Colombia¹ and is also known as Bobonaza Quichua (Orr & Wrisley 1981) and Pastaza Quichua (Orr 1991), since most speakers live along the Bobonaza river in the Pastaza province of Ecuador (Rice 2020: 182).

In Ecuador, the varieties of Quechua are referred to as “Kichwa/Quichua” and, traditionally, all varieties were considered to be dialects of a single Kichwa/Quichua language. However, recent research has moved towards positing at least two languages: Ecuadorian Highland Kichwa, which is spoken in the Ecuadorian Andes, and Amazonian Kichwa,² which is spoken primarily in the Ecuadorian Amazon with some extensions into the Peruvian Amazon (Grzech et al. 2019; Muysken 2019). Under this view, Northern Pastaza Kichwa is thus a dialect or variety of Amazonian Kichwa.

Estimates of the number of speakers of Northern Pastaza Kichwa range between 10,000 and 30,000 and the vitality status of Northern Pastaza Kichwa has been rated as “threatened” and “definitely endangered” (Rice 2020: 186). Despite the number of speakers, Northern Pastaza Kichwa and the other Indigenous languages of Ecuador face the challenge of stagnating intergenerational transmission. It has become increasingly difficult to find fluent speakers under the age of 30 and young adults and youth appear to feel more confident speaking Spanish (Grzech et al. 2019: 131–132; Rice 2020: 186).

¹This branch of Quechua is also sometimes referred to as “Late Northern Quechua” (Muysken 2021). This branch is listed on Glottolog as “Colombia-Ecuador Quechua”: <https://glottolog.org/resource/languoid/id/colo1257>.

²Amazonian Kichwa is sometimes referred to as “Ecuadorian Lowland Quechua” in some descriptive literature.

In terms of typological characteristics, Northern Pastaza Kichwa looks like other Quechuan languages: small vowel inventory, agglutinative word structure, exclusively suffixing, and heavy use of verbs and nominalizations. That said, Northern Pastaza Kichwa and the other varieties in the northern branch of Quechua are somewhat unique in that they are morphosyntactically “simplified” compared to the rest of the Quechuan family. Muysken has argued extensively that the Ecuadorian-Colombian Quechua varieties represent “pidginized” or “creolized” forms of Quechua that arose from the demographic collapse and reshuffling of Indigenous populations in Ecuador following contact with European colonizers in the sixteenth century. Speakers of non-Quechuan languages would have then shifted to one or more varieties of Quechua as the available lingua franca after their communities were destroyed by disease and colonial violence, thus resulting in the phonological and morphosyntactic simplification of the Quechuan varieties spoken there (Muysken 2000, 2009, 2019, 2021).

“Kichwa/Quichua” is an exonym that is used most frequently in academic literature and socio-political discourse in Ecuador. Northern Pastaza Kichwa speakers refer to their language as *runa shimi* ‘human language’ and themselves as *runa* ‘human’. There is no single ethnic identity associated with Northern Pastaza Kichwa. Some speakers identify themselves as belonging to one of the officially recognized Indigenous nationalities of Ecuador: Andwa, Sápara, or Kichwa Amazónico. Other speakers eschew the national identities and instead self-identify on the basis of their community of origin. Northern Pastaza Kichwa is also the second or third language of speakers of Wao Tedeo and the Chicham languages,³ and intermarriage between the language groups is common (Reeve 2022: 5–6). Traditionally, Northern Pastaza Kichwa speakers practiced subsistence agriculture, hunting, fishing, and foraging in the rainforest as well as riverine trade with extended kinship networks. Today, this lifestyle is becoming increasingly non-viable due to political, economic, and environmental pressures. Many families are forced to abandon their lands and communities and move to cities and towns to participate in wage-labour employment. The resulting erosion of traditional material culture and lifestyle further accelerates language shift to Spanish.

2.2 Corpus and participants

The data used for this chapter are drawn from video recordings collected between 2011 and 2022 by multiple researchers. The majority of the recordings in the cor-

³Wao Tedeo [ISO 639-3: auc] is an isolate also known as Wao Terero, Wao, and Huaorani. The main Chicham language spoken in this area of Ecuador is Achuar-Shiwiar [ISO 639-3: acu].

pus were collected by Tod Swanson and Janis Nuckolls. Some of the videos are publicly accessible on Swanson's (2024) and Nuckolls's (2024) YouTube channels. Other recordings come from my own fieldwork, and Lisa Warren Carney. Each recording has a unique file name which combines the collecting researcher's initials, the language's ISO code, and a three-digit numeric ID. Recordings beginning with "tds_qvz" come from Tod Swanson, "jbn_qvz" from Janis Nuckolls, "lwc_qvz" from Lisa Warren Carney. My own recordings are not marked with my initials and read simply as "qvz_001", "qvz_002", etc. I manage the corpus with SayMore (2022) and use FLEX (2021) and ELAN (2024) for annotation.

The corpus is comprised of approximately six hours of video recordings and 30,710 words. All of the videos have been transcribed and translated into Spanish and/or English. The transcriptions and translations were completed by native speakers with some assistance from researchers. The genres represented in the corpus are principally monologues and conversation, with some song, and procedural description. The recordings come from seven consultants, all adults, one male and six female, the youngest of whom at the time of recording were in their mid 30's and the oldest in their 80's.

The corpus is not balanced in terms of speaker gender since most of the participants recorded are older women and all but two of the speakers are related to each other. Some speakers are also better represented than others in terms of recordings. The corpus is thus an example of convenience sampling, in which there was no conscious selection of recordings; the recordings that are available and ultimately used are a matter of coincidence rather than design (Seifart 2008: 64).⁴ A summary of the recordings used in the corpus is given in Appendix A.

2.3 Fillers: Hesitatives and placeholders

Fillers are non-silent linguistic devices used to buy time while experiencing a disfluency event and retrieving a sought after lexical item (Fox 2010). The two kinds of fillers relevant to this chapter are hesitatives and placeholders. The two categories of filler are distinguished by their degree of syntactic integration with the surrounding utterance.

Also known as "filled pauses" (Lickley 2015), hesitatives are used by speakers to signal to the interlocutor that they are experiencing a disfluency. Hesitatives are used to announce that the speaker is searching for a word or is wanting to

⁴Convenience sampling is often the case when working with Indigenous/minority language data. Language communities engaged in a documentation project may wish to prioritize collecting recordings of elders or specific genres of language use such as traditional narratives and discourse associated with traditional material culture.

keep or cede the floor (Clark & Fox Tree 2002). Hesitatives can be as simple as an unspecific speech sound, such a drawn-out vowel or nasal sound, and more conventionalized but not necessarily lexical forms such the English hesitatives *uh* and *um*. Lexical forms and even phrasal items can also be deployed as hesitatives, such as the English *well*, *like*, and *y'know* (Fox 2010: 1).

Placeholders can also be used when a speaker experiences a disfluency while searching for a particular lexical item. However, unlike hesitatives, placeholders exhibit more morphosyntactic integration with the adjacent/surrounding utterance (Hayashi & Yoon 2006: 490). A placeholder thus, “holds the place” of the targeted lexical item that the speaker experiences difficulty accessing. The posterchild of (hesitative) placeholders in English is *whatchamacallit* which derives from something like ‘what you may call it’. An English speaker may use *whatchamacallit* during a disfluency event to hold the place of a lexical item the speaker is having difficulty acquiring.

Not all placeholders are used in contexts of hesitation or disfluency. Even if the target word or phrase is accessible to the speaker’s short term memory, placeholders can be deployed euphemistically to avoid uttering taboo words or topics (Enfield 2003), or to purposefully obfuscate referents or propositions from certain interlocutors (Amiridze 2010). Placeholders can also be deployed when the target – even if accessible to the speaker – simply is not important to mention and, thus, results in purposefully vague or imprecise expressions (Palacios Martínez & Núñez Pertejo 2015; Podlesskaya 2010: 26; Seraku 2022).

The majority of placeholder uses of *mashfi* in this corpus are used in contexts of disfluency ($N = 17$), while a handful of the tokens are used for in the vague/ imprecise referential manner ($N = 5$) for which the target is never uttered. Both types of placeholders are discussed in §3.2, but generally this chapter is given to the discussion of *mashfi* placeholders for which there is an identifiable target.

In discourse, placeholders uttered in contexts of disfluency are typically followed by utterances containing the target word when the speaker resolves the disfluency (Keevallik 2010: 310; Vallejos Yopán 2023: 12). In Northern Pastaza Kichwa, placeholders that target a delayed constituent tend to carry the same morphology and syntactic position as the target item. Podlesskaya (2010: 18) refers to this phenomenon as “mirroring” the “grammatical shaping” of the target and notes that placeholders can fully or partially mirror the target.

Cross linguistically, common lexical sources of fillers are demonstratives, pronouns, semantically bleached nouns, and lexicalized constructions featuring interrogatives (Podlesskaya 2010: 12–13). In Northern Pastaza Kichwa, demonstrative pronouns are often deployed as fillers; however, other lexical items like dis-

course particles and conjunctives borrowed from Spanish can also be utilised as fillers. Table 1 shows the lexical items commonly used as fillers in the language.

Table 1: Lexical items used as hesitative fillers in Northern Pastaza Kichwa

Word	Gloss
<i>chay</i>	mirative distal demonstrative pronoun
<i>chi</i>	distal demonstrative pronoun
<i>este</i>	filler (borrowing from Spanish)
<i>ima</i>	interrogative pronoun ‘what’
<i>mashti</i>	hesitative; placeholder; resumptive pro-verb
<i>kay</i>	proximal demonstrative pronoun
<i>ñā</i>	temporal discourse particle ‘now, then, already’
<i>osea</i>	conjunctive (borrowing from Spanish)
<i>y</i>	conjunctive (borrowing from Spanish)
<i>ya</i>	now, already (borrowing from Spanish)

Most of the lexical items in Table 1 are native Quechuan roots and can be used as hesitatives. The demonstrative pronouns are among the most frequently occurring words in the corpus, and I have not yet determined how often they are used as hesitatives nor if any of them can be used as placeholders. Northern Pastaza Kichwa speakers also make use of non-lexical hesitatives. The example in (1) shows the use of the lengthened /a/ vowel as a hesitative.

- (1) *aaa chi runduma-ta chi-ta mana shuti-ta yacha-ni=chu*
HES DIST.DEM sedge-OBJ DIST.DEM-OBJ NEG name-OBJ know-1=NEG
‘um... that sedge, that one, I don’t know its name’ [qvz014:22]

The other fillers given in Table 1 are borrowings from Spanish and have more varied lexical functions when not deployed as hesitatives. The borrowing of the Spanish demonstrative *este* is used only as a hesitative in Northern Pastaza Kichwa discourse. The use of *este* as a filler is quite common in Latin American Spanish (Kany 1969; Vallejos Yopán 2023; Zorraquino et al. 1999).

Compared to the other fillers in Northern Pastaza Kichwa, *mashti* has some unique properties: it likely derives from a lexical phrase (cf. §5) instead of a demonstrative, it can be deployed both as hesitative and a placeholder and can be used as a resumptive pro-verb (cf. §3.1). Because of these unique properties, I chose to focus on *mashti* for this chapter. I identified 157 *mashti* tokens in the

corpus and annotated each for speaker, function (hesitative, placeholder, or proverb), word class (if placeholder), target (if placeholder or proverb), direction to target (before or after), and the disfluency correlates described in §4. This annotation data can be found in the `mashiti_dat.csv` file, which is available in an OSF drive (A. Rice 2023).⁵ Likewise, the video, audio, and ELAN files for the recordings that the tokens were drawn from are also available in the OSF drive. The counts presented in the tables in this chapter were generated by an R script (2024) and plots were generated using the `ggplot2` package for R (Wickham 2016). The script is available in the aforementioned drive.

2.4 The multimodal expression of disfluency events

In this chapter, I take a multimodal approach to the use of fillers in that I consider disfluency phenomena in two modalities: the vocal-aural modality and the visuo-spatial modality.⁶ The vocal-aural modality concerns the production, reception, and interpretation of speech as an acoustic signal. The visuo-spatial modality refers to the use of bodily articulators beyond the vocal tract that contribute to meaning in language (posture, facial expression, hand gestures, direction of gaze, etc.). Table 2 shows the disfluency correlates that I investigate in this chapter and their corresponding modalities.

Table 2: Disfluency correlates by modality

Modality	Disfluency correlate
Vocal-aural modality	false starts
	pauses
	prosodic lengthening
Visuo-spatial modality	gaze aversion
	excessive blinking
	manual gesture
	gestural disfluency

Many of the examples in this chapter are presented in a format to help convey the visually perceived bodily elements that occur with the prosodic and lexical material. In the examples in this chapter, I make use of special conventions to

⁵<https://osf.io/g84mv/>.

⁶Cf. Stivers & Sidnell (2005) for a detailed primer on multimodality and the distinction between the vocal-aural and visuo-spatial modalities.

indicate different types of disfluency correlates. These conventions are given in Table 3.

Table 3: Disfluency conventions used in the examples in this chapter

Disfluency correlate	Convention
false start	<i>word-</i>
pause	(pause—XXXms)
prosodic lengthening	<i>worrrd</i> (XXXms)
duration and direction of gaze	=>(right), <=(left), =^ (up), =^ (down)
excessive blinking	^
duration of manual gesture	***
duration of gesture hold	(hold—XXXms)
number of repetitions in stalling gestures	(X repetitions)

False starts are indicated by “-” and are attached to the word that was started in error in the transcription line. The word may be fully or partially produced. False starts are represented as [false.start] in the gloss line. Pause indicators are placed in parentheses in the transcription line and are followed by an em dash with the length of the pause in milliseconds. Words that are prosodically lengthened are represented in the transcription line by repetition of the letter that corresponds to the lengthened segment, followed by a set of parentheses containing the duration of the lengthened segment in milliseconds.

The conventions used to represent disfluency correlates in the visuo-spatial channel are placed above the transcription line. The use of “=” indicates an abrupt shift in gaze and repeated instances of “=” are used to indicate the approximate duration of the new gaze direction relative to the utterance in the transcription line. The use of the caret symbols (>, <, ^, ^) are appended to each gaze annotation to indicate the direction of gaze (from the point of view of the camera). Excessive blinking events are indicated by repeated circumflex symbols (^) and each circumflex represents one blink. The use of repeated asterisks (*) shows the duration of a manual gesture.⁷ Indicators for gestural holds are placed in parentheses following the asterisk strings for manual gestures. For stalling gestures, the number of repetitions is given in parentheses following the gesture indicator.

⁷The convention of using “***” to indicate the relative duration of a gesture is an adaptation of Kendon’s (2004) convention of using “***” above the transcription to show the duration of a gestural stroke.

Many of the examples in this chapter represent longer stretches of discourse and, accordingly, examples are divided into intonational units using lowercase letters (a, b, c, etc.). The examples that show disfluency correlate in the visuo-spatial channel are accompanied by a figure that shows one or more still images from the video recording that the *mashfi* token was drawn from. The figure caption contains the recording name and the start and stop time stamps. If more than one still is present in the figure, each still is indexed with a lowercase roman numeral (i, ii, iii, etc.). In the associated example, the index is placed above the transcription lines to show when the still occurs relative to the production of the utterance.

To illustrate the use of some of these conventions, consider Figure 1 and the associated stretch of discourse in example (2). In (2a), the speaker is narrating a story. In (2b), she begins a new intonational unit and the index (i) above the transcription refers to still (i) of Figure 1, wherein the speaker is shown to be looking down at her hands while she is narrating. In (2c), the speaker begins another intonational unit but experiences a disfluency. The discourse particle *ña* (in bold) is used as a hesitative; this is followed by a pause of 1456ms, an abrupt change in gaze to the right (the speaker's left), and a period of excessive blinking towards the end of the pause. The index (ii) above the transcription line of (2c) refers to still (ii) of Figure 1, which captures the moment the speaker shifts her gaze to the right (as is indicated by the red arrow).

In (2d), the use of “<==” shows that speaker shifts her gaze to the left (her right). She also swings her right arm out to her right, extends her pointer finger, and brings it into contact with the table she is sitting at. The use of the asterisks (****) show the approximate duration of this manual gesture. The index (iii) in (2d), points to (iii) of Figure 1, wherein the same actions are indicated with the red arrows. After starting the gesture and gaze shift, the speaker uses the distal demonstrative pronoun *chi* as a hesitative. The production of *chi* is prosodically lengthened; the extra {i} characters show that the final /i/ vowel of *chi* was drawn out, and the duration of the lengthened /i/ is given in the following parentheses (311ms). This is followed by a pause of 639ms. In (2e), the disfluency is resolved, and the speaker continues the story, although her gaze remains directed towards the left (not shown in Figure 1).

- (2) a. *peruwano-y maska-nга ri-sha tupa-nau-ra chi libru-ta*
Peru-LOC search-FUT go-ss find-3PL-PST DIST.DEM book-OBJ
'they had gone searching in Peru [and] they found the book'

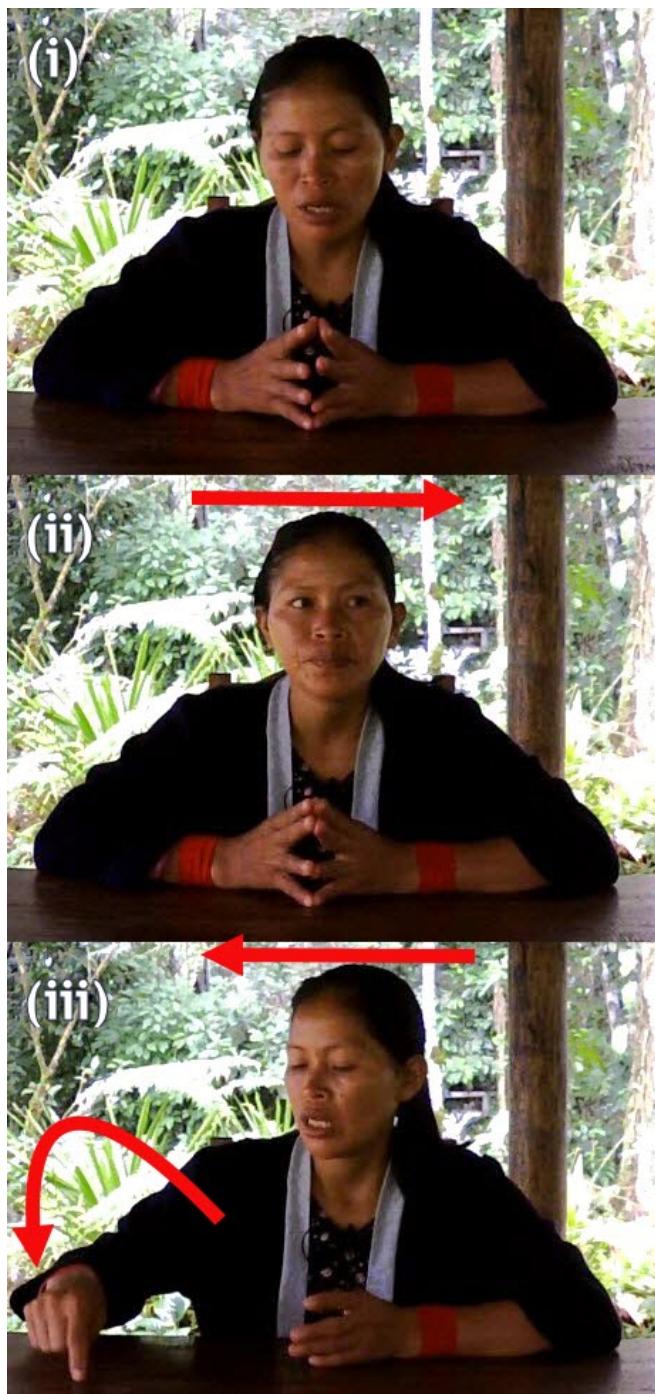


Figure 1: Gaze shift and gesture in a disfluency event (qvz026, 7:08–7:18), photo © 2019 Alexander Rice.

- b. (i)
chi-wi
DIST.DEM-LOC
'there'
- c. (ii)
 ^ ^ ^
 =====>
 ña (pause—1456ms)
DSC.PRT
'then...'
- d. (iii)

=====
- chiii*(311ms) (pause—639ms)
DIST.DEM
'thaat...'
- e. =====
 saki-ri-k *runa*
leave-REFL-SBJ.NMLZ human
'person who had stayed behind' [qvz026:140–144]

Figure 1 and (2) provide an example of what disfluency events look like in Northern Pastaza Kichwa in both the vocal-aural and visuo-spatial modalities. The speaker used two lexical items that are commonly deployed as fillers in the language: the demonstrative pronoun *chi* and the discourse particle *ña* (cf. Table 1 in §2.3). The deployment of these lexical items as hesitative fillers is made obvious by their association with correlates of disfluency. In §4 I discuss these correlates of disfluency in greater detail using multimodal examples like the one shown here.

3 Filler and pro-verb functions of *mashti* in Northern Pastaza Kichwa

In this section, I give an overview of the three different discourse functions of *mashti* as represented in the corpus. Of the 157 total tokens in the corpus, the

Table 4: *Mashti* tokens by type in corpus

<i>mashti</i> type	Count
hesitative filler	119
placeholder filler	22
pro-verb	16

majority (75%) are used as hesitative fillers. Table 4 gives the total counts of *mashti* tokens by type.

Figure 2 shows the counts of *mashti* tokens by type and speaker. The speakers in Figure 2 are ordered by age at time of recording, from mid 30s (SV) to 80s (LC). Hesitative filler uses of *mashti* (H) appear to be the most frequent use of *mashti* for most speakers. However, as discussed in §2.2, the corpus is not balanced in terms of recordings per speaker. For instance, speaker BD is overrepresented in this corpus compared to the others. Nevertheless, most speakers appear to use the hesitative filler *mashti* more than the placeholder filler or pro-verb uses of *mashti*.

I define the different types of *mashti* by functional and referential criteria which tend to be reflected in the forms of *mashti*. Filler instances of *mashti* are distinct from pro-verb uses of *mashti* in that fillers are used by the speaker to signal that they are experiencing disfluency. Pro-verb instances of *mashti*, by contrast, do not signal hesitation and their principal function is to point to previously uttered propositions and events. The two types of filler *mashti* – hesitations and placeholders – are distinguished by whether or not they are syntactically integrated into the surrounding discourse. Hesitations are not integrated, while placeholders are and, consequently, do not have additional morphology while placeholder uses of *mashti* do.

These functional and distributional generalizations are useful for categorizing most of the *mashti* tokens in the corpus. However, there are some tokens that would seem to straddle the categories as set up here. In particular, there are two “undetermined” instances of *mashti* in the corpus. I discuss these two instances in §3.1 along with the hesitative uses of *mashti*. Placeholder and pro-verb uses of *mashti* are discussed in §3.2 and §3.3, respectively. In §3.4 I consider the variation in the base form of *mashti*, also produced as *imashti*

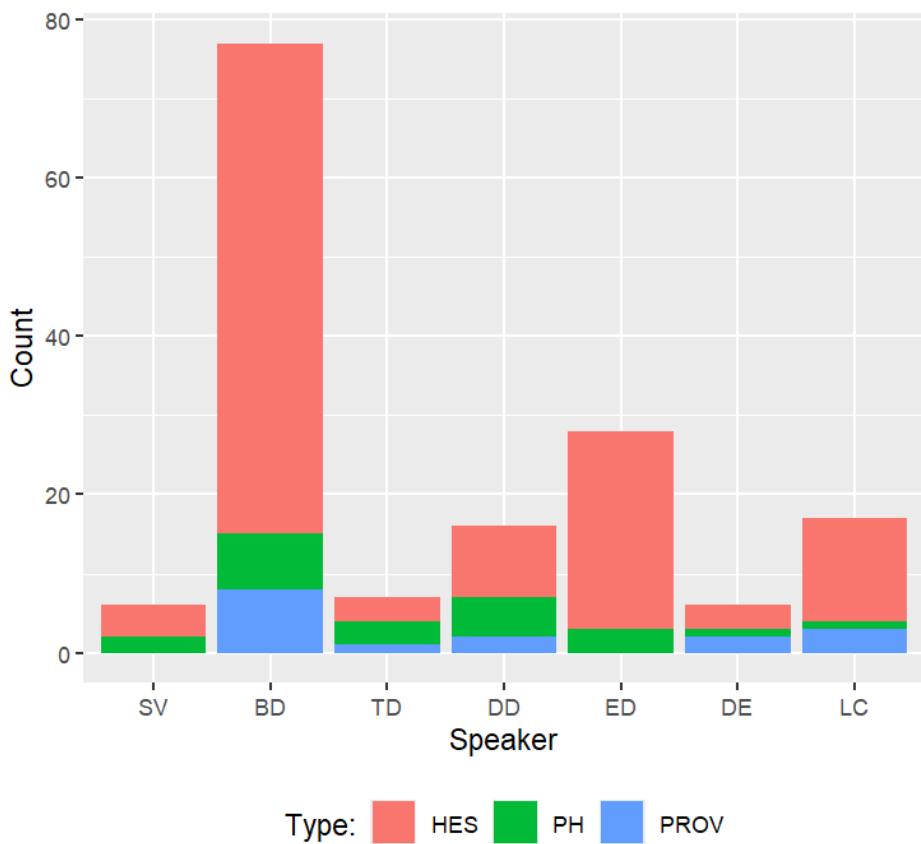


Figure 2: Counts of *mashti* type by speaker

3.1 Hesitative filler *mashti*

As a hesitative, *mashti* signals that the speaker is experiencing disfluency and serves to buy time while the speaker decides what to say next. The hesitative is not syntactically integrated into the surrounding utterance (Hayashi & Yoon 2010: 43). The example in (3) comes from an elicitation session I had with a consultant, wherein I showed her a cartoon animation and asked her to recount the sequence of events. In (3a), she uses *imashti* as a hesitative to buy time while she organizes her thoughts and prepares to recount the events of the cartoon animation.

- (3) a. *kay pelikula-y riku-shka-nchi imashtiii* (pause—1690ms)
 PROX.DEM movie-LOC see-PERF-1PL HES(475ms)
 ‘in this movie we have seen ummm...’
- b. *misi-wan alku-wan* (pause—1536ms)
 cat-INST dog-INST
 ‘a cat with a dog’
- c. *chiii* (pause—1372ms)
 DIST.DEM(804ms)
 ‘thaat....’
- d. *misi allku-ta wata-sha*
 cat dog-OBJ tie.up-ss
 ‘the cat [was] tying up the dog’ [qvz010:22–25]

The *imashti* in (3a) does not occupy any syntactic slot in the utterance due to the distributional and morphological tendencies of lexical items in Quechuan languages. Like all other Quechuan languages, constituent order in Northern Pastaza Kichwa is flexible but tends towards SOV and modifiers like adjectives and adverbs precede their respective heads. Placed after the verb *rikushkanchi* ‘we have seen’, *imashti* is not acting as the subject, object, or modifier of any of the other constituents. Even if assuming non-standard word order, we can rule out the use of *imashti* in (3a) as a placeholder targeting an object on a morphological basis because objects in Quechuan languages – whether direct or indirect – are obligatorily marked with case suffixes.

There is also a prosodic case to be made for the hesitative classification of *imashti* in (3a). It appears on the boundary of the intonational unit, and the production of the word itself is lengthened. Specifically, the final vowel is prolonged for 475ms. This is followed by a 1690ms pause before the next intonational unit.⁸ Prosodic differences between the different uses of *mashti* are discussed in more detail in §4.1, but prosodic lengthening is likely a key indicator of hesitative fillers compared to placeholder fillers.

A hesitative *mashti* can be more prosodically integrated into a phrase without being syntactically integrated. In (4), *mashti* does not occur at either boundary of the intonational unit, it is not prosodically lengthened, and there are no associated pauses. Nevertheless, it is not syntactically integrated into the utterance.

⁸Note also the disfluency episode in (3a-b) is replicated in (3c-d). In (3c) the distal demonstrative *chi* is used a hesitative and follows the same pattern as the *imashti* of (3a). It is prosodically lengthened and followed by a pause before the target is acquired in (3d).

The *mashti* “interrupts” the syntactic relationship between the proximal demonstrative *kay* that points to the noun *muyuwallata* ‘little fruit’. If this *mashti* were a placeholder targeting the fruit, it would bear at least some of the same suffixes that *muyu* ‘fruit’ bears. Thus, the *mashti* in (4) does not factor into the utterance’s syntax.

- (4) *kay mashti muyu-wa=lla-ta apa-sha shamu-ra-ni*
PROX.DEM HES fruit-DIM=LIM-OBJ take-ss come-PST-1sg
ni-shka-∅=lla
say-PERF-3SG=LIM
“I came taking this, um, just this little fruit” he had just said’
[tds_qvz036:89]

The presence or absence of morphology on a given instance of *mashti* is, thus, a rule of thumb that can be used to distinguish the hesitative *mashti* from the placeholder or proverb uses of *mashti* if morphology is expected on the target. However, there are instances in which the degree of syntactic integration of *mashti* is not clear. In such cases, additional contextual criteria may be more useful.

The following example shows one of the two *mashti* tokens in the corpus which I defined as hesitative despite having additional morphology. The example in (5) and the associated Figure 3 come from a recorded conversation I had a with speaker, wherein she taught me about various medicinal plants. In (5a), the speaker describes how the roots of a type of wild ginger can be processed to create a kind of shampoo. She ends the intonational unit in (5a) with *imashtiga*, which bears the topic enclitic *=ga*. Apart from occurring at the boundary of the intonational unit, there are no other indicators of prosodic separation, such as lengthening or pauses. It is also not clear to which extent *imashti* is syntactically integrated. The inclusion of the *=ga* topic enclitic may indicate that *imashti* is actually a placeholder that targets *kaygunaga* ‘these ones’ in (5b), which also carries the topic enclitic. If so, this would be a case of partial mirroring instead of full mirroring (cf. §3.2), because *imashti* does not carry the plural suffix *-guna* that the target demonstrative pronoun has.

- (5) a. (i)
chi-guna m=a-u-n sapi pay sapi-wan imashti=ga
DIST.DEM-PL EV.S=COP-PROG-3 root 3 root-INST ?=TOP
‘they’re those ones, and [you mash] its root, [you prepare the
shampoo] with its root, um’



Figure 3: A speaker hesitates then points to a plant (qvz015, 9:38–9:44),
photo © 2019 Alexander Rice.

b. (ii)

 kay-guna=ga panga panga
 PROX.DEM-PL=TOP leaf leaf
 ‘these ones, [that] leaf, [that] leaf’

c. *chi=was panga m=a-naun*
 DIST.DEM=MIR leaf EV.S=COP-3PL
 ‘[and] that leaf, those are the ones’ [qvz015:133–135]

The use of *imashti* in (5a) could also be interpreted as that of a hesitative filler. Evidence for this argument comes from the speaker’s gesture. In (5a) and (i) of Figure 3, the speaker has their hands together after performing a gesture mimicking the mashing movement to make the shampoo from the ginger roots. Her hands remain in this position while she utters *imashti* in (5a). In (5b) and (ii) of Figure 3, she points to another cluster of ginger plants while uttering *kaygunaga* ‘these ones’. Had she performed the gesture concurrently with, or closer to, her production of *imashti* in (5a), that would have made a stronger case for this *imashti* being a placeholder, because the gesture would clearly indicate the target of the placeholder. I lean towards interpreting the use of *imashti* in (5a) as a hesitative filler and represented it as “um” in the translation line.

The other use of *mash* that I determined to be a hesitative filler despite the presence of additional morphology is given in (6), which comes from the preamble of a traditional narrative. The speaker is talking about how her grandparents

would tell stories to her and her siblings when she was a child. In (6), the speaker uses a *mashti* that is prosodically integrated into the utterance, but there is no obvious syntactic relation it might have to other constituents.

- (6) *chi-manda apamama mashti=shi chasna chi-guna-ta*
DIST.DEM-ABL grandmother ?=EV.O like.that DIST.DEM-PL-OBJ
kwinta-k a-shka-una apamama
tell-SBJ.NMLZ COP-PERF-3PL grandmother
'from there, grandmother [and grandfather] um, like that would tell those
ones [stories]' [tds_qvz044:33]

Curiously, this instance of *mashti* also carries the evidential "other perspective" enclitic *=shi*, which is used to mark information as coming from a perspective other than the speaker (Nuckolls 2018).⁹ There are no other constituents with the same enclitic in the utterance or the following intonational unit – which is not shown in (6) – making this *mashti* unlikely to be a placeholder. At the same time, it is not obvious what the addition of an evidential enclitic would contribute to a hesitative filler. The native speaker who translated this narrative represented this *mashti* in her Spanish translation as *este*, which is a Spanish demonstrative that is commonly deployed as a filler in Latin American and Amazonian Spanish (Vallejos Yopán 2023). Beyond that, the translator did not know what to make of *mashtishi* in (6). It is possible that the inclusion of *=shi* is a speech error.

3.2 Placeholder *mashti*

Placeholder fillers are used in disfluency events as a substitute for a specific lexical item (the target), which is later produced when the disfluency is resolved. Unlike hesitative fillers, placeholder fillers are syntactically integrated into their surrounding utterances (Hayashi & Yoon 2006: 490). Table 5 shows the counts of the different targets of the placeholder uses of *mashti*. The five independent placeholders that do not have a target are discussed below.

The placeholder filler use of *mashti* is exemplified in (7), wherein the consultant is telling an "Amazonian" version of the biblical Noah's Ark story. The characters in the story are building the ark by splitting the trunk of a *tarapoto* tree. The consultant experiences a disfluency in recalling the name of a tree and uses *mashti* in (7a) as a substitute before acquiring and producing the name of the tree, which – as the target – is underlined in (7b).

⁹Evidentials in Quechuan languages are often used for non-evidential functions (Bendezú-Araujo 2023; Grzech 2016b; Nuckolls 2014).

Table 5: Target types of placeholder *mashti* ($N=22$)

Target type	Count
adposition	1
converb	3
deverbal noun	1
none (independent)	5
noun	11
finite verb	1

- (7) a. *kay mashti-taaa* (pause—664ms)
 PROX.DEM PH-OBJ(122ms)
 ‘this whatchamacallit’
- b. *tarapu-ta wakta-sha*
 tarapoto-OBJ hit-ss
 ‘[this] tarapoto [tree], [they were] hitting [it]’ [tds_qvz044:59–60]

This use of *mashti* is not prosodically integrated into the utterance, as shown by the prosodic lengthening and the introduction of a pause that splits the verbal construction between two intonational units. However, this *mashti* is syntactically integrated into the utterance. The target of the *mashti* placeholder – the *tarapoto* tree, – is the object complement of the *wakta-* ‘hit’ verb in (5b) and thus carries the object suffix *-ta*. The placeholder *mashti* in (5a) is likewise suffixed with the object suffix *-ta*, and the fact that *mashti* is suffixed with the same case marker is an indicator of the syntactic integration of *mashti* in the utterance as the object complement of *waktasha* ‘hitting’.

The phenomenon of a placeholder matching the grammatical shaping of the target word has been referred to as “mirroring” (Podlesskaya 2010: 18). A placeholder filler can partially or fully mirror the eventually produced target word. The following example in (8) shows a placeholder *mashti* that fully mirrors the target lexical item. The *mashti* in (8b) includes not only the case suffix, but also the evidential enclitic *=shi*.

- (8) a. *chiga riku-kpi=ga*
 DSC.PRT see-DS=TOP
 ‘and then seeing’

- b. *imashti-ta=shiii*
PH-OBJ=EV.O(148ms)
‘an apparent whatchamacallit’
- c. *ima* *yutu-ta=shi* *intiru-ta yanu-shka-ta*
what[false start] partridge-OBJ=EV.O whole-OBJ boil-PERF.NMLZ-OBJ
‘a- an apparent partridge, a whole one, a boiled one’ [qvz007:241–243]

The following example in (9) shows instances of full and partial mirroring of *mashti* placeholders. The first set consists of *imashti* with the instrumental suffix *-wan*: it fully mirrors the target noun *sapi* ‘root’ that also bears the same instrumental suffix. The second *mashti* placeholder *mashtigunaga* partially matches the target nominal by bearing the same plural suffix but has a topic enclitic rather than the emphatic enclitic that is found on the target noun.

- (9) *mashti pay-ba imashti-wan pay-ba sapi-wan mashti-guna=ga*
HES 3-GEN PH-INST 3-GEN root-INST PH-PL=TOP
runduma-guna=ya
sedge-PL=EMP
‘um, with its whatchamacallit, with its root [of the] whatchamacallits,
[of the] sedges!’ [qvz015:131]

The examples thus far show the placeholder *mashti* targeting object nouns. The placeholder can also target other material like converbs and deverbal nouns as shown in (10). In (10b), the *mashti* placeholder bears the same-subject suffix *-sha*, which is suffixed to verbal roots that are used as converbs that head adverbial clauses. The *-sha* same-subject suffix indicates the subject of the adverbial clause is the same as that of the following clause. The placeholder *mashtisha* in (10b) mirrors its target *lluchusha* ‘peeling’ in (10c). Note also that the entire adverbial clause that stretches between (10a) and (10b) *kayba karata mashtisha* ‘whatchamacalliting this one’s bark’ is reproduced with the target lexical item in (10c) *kayba karata lluchusha* ‘peeling this one’s bark’. Not only was the target word eventually produced, but the entire clause is repeated with the correct target word.¹⁰ In (10d) *mashti* is used as a placeholder with the infinitive nominalizer *-na* to hold the place for the deverbal noun *raspana* ‘to grate’.

- (10) a. *chiga chay-ta kay*
DSC.PRT DIST.DEM-OBJ[false start] PROX.DEM[false start]

¹⁰Podlesskaya (2010: 23) refers to this phenomenon as “recycling”.

- kay-ba kara-ta*
 PROX.DEM-GEN bark-OBJ
 ‘and so, that one- this- this one’s bark’
- b. *mash**ti-shaaa*** (pause—997ms)
PH-ss(190ms)
 ‘whatchamacalliting’
 - c. *eee kay-ba kara-ta lluchu-sha*
 HES PROX.DEM-GEN bark-OBJ peel-ss
apa-sha=mi (pause—932ms)
 take-ss=EV.S
 ‘um... this one’s bark, peeling [it] and taking [it]
 - d. *eee mas**hti-naaa** a-n raspa-na*
 HES **PH-INF.NMLZ** (118ms) COP-3SG grate-INF.NMLZ
 ‘um... to whatchamallit, to grate [it]’ [tds_qvz033:133–136]

The example in (10) also shows several overt signifiers of disfluency. Both uses of the *mashti* placeholders are prosodically lengthened and occur near pauses. In (10a), there are two false starts, and non-lexical hesitative fillers are used in (10c) and (10d).

The previous examples show instances of *mashti* being deployed as a placeholder filler, or a placeholder used in contexts of hesitation. However, not all placeholders are used in contexts of hesitation. Placeholders can be deployed for a variety of discourse functions (Enfield 2003). Of the 22 placeholder uses of *mashti* in the corpus, 18 are deployed as fillers in contexts of hesitation that have specific targets that are later produced in the same or a following utterance. The other four are used as “vague” or “imprecise” expressions that have no expressed target in discourse. These placeholders are used when the targeted entity or proposition is known to the speaker and interlocutor in the context of discourse and specific reference to that entity or proposition is not deemed necessary by the speaker (Podlesskaya 2010: 26). I refer to these as “independent” placeholders. In (11), an independent placeholder is used during a speaker’s account of an experience from her life before she was married. The speaker recalls that her grandmother had accused the speaker and her cousins as being lazy and improper for not yet having husbands. The grandmother’s discourse is represented by the speaker in (11a)-(11c).

- (11) a. *warmi-wa-guna kasna kausa-nga ra-u-ngichi ni-sha=s*
woman-DIM-PL like.this live-FUT do-PROG-3PL say-ss=CONJ
'and [she's] saying: "[as] young women you will live like this"
- b. *kungay=lla*
peace=LIM
'just peacefully'
- c. *warmi m=a-ngichi yacha-na m=a-u-ngichi*
woman EV.S=COP-2PL know-INF.NMLZ EV.S=COP-PROG-2PL
imasna mashti-na-ta=s ni-sha=s
how.many PH-INF.NMLZ-OBJ=CONJ say-ss=CONJ
'she's saying: "you all are women, you all need to know [how to do]
all kinds of stuff"
- d. *ñña chasna a-shka-y-bi ñuka mama-ma*
DSC.PRT like.that COP-PERF.NMLZ-LOC-LOC 1 mother-ALL
kuti=lla-ta bulltiya-ni
again=LIM-ADV return-1SG
'so after it had been like that, I return again to my mother'
[qvz042:145–148]

In (11c), a placeholder *mashti* occurs as head of the object complement of *yachana maungichi* 'you all need to know'.¹¹ The object marker indicates that this use of *mashti* is syntactically integrated into the utterance. However, there is no target for this *mashti*, neither before nor after (11c). Explicit reference to the target of this *mashti* is not needed, because the context makes it clear that *mashti-natas* refers to domestic chores that are often expected of women in Amazonian Ecuador.

In most cases, placeholder uses of *mashti* are easy to detect because of the presence of morphology indicating syntactic integration and a following target in the case of placeholder fillers. However, placeholder fillers that target non-derived nouns in subject positions are more difficult to detect because there is no case marking or any kind of obligatory morphology that occur on subject nouns in Northern Pastaza Kichwa and in all other Quechuan languages (Adelaar & Muysken 2004: 213). Because of this, distinguishing between a hesitative filler

¹¹In this example, the object follows the verb, which is a non-standard but acceptable ordering of constituents in Northern Pastaza Kichwa.

mashiti and a placeholder filler *mashiti* that targets a subject noun can be difficult, as is shown in (12).

- (12) a. *aysa-shka m=a-ra-∅ yaya shuk saltakama*
 pull-PERF.NMLZ EV.S=COP-PST-3SG father one armoured.catfish
 ‘one armoured catfish was pulled [caught] by my father’
- b. *ishkay mashitiⁱⁱ muta*
 two **HES/PH**(74ms) muta.catfish
 ‘[and then] two, um/whatchamacallits, muta catfish’
 [jbn_qvz002:297–295]

In (12a) the speaker lists the fish her father caught one day. In (12b) she experiences disfluency and forgets the name of species of the two catfish that her father had caught. After the numeric determiner *ishkay* ‘two’ she uses *mashiti*. This is followed by the target *muta* – a type of catfish – which does not carry any additional morphology and, thus, this *mashiti* could be a placeholder that targets, and fully mirrors, *muta*. The numeric determiner could then be modifying *mashiti* to produce an utterance akin to; “two whatchamacallitss, *muta* catfish” On the other hand, this *mashiti* may be a hesitative filler that does not participate in the utterance’s syntax. Prosodically, the *mashiti* in (12b) is lengthened, which would appear to be more associated with hesitative fillers compared to placeholders, but 74ms only barely passes the threshold I establish for lengthened segments in this phonological context (cf. §4.1.3).

My rule of thumb in which a *mashiti* lacking additional morphology is likely a hesitative filler may thus overcount hesitative filler instances of *mashiti* at the expense of placeholder filler uses of *mashiti*. This is an imperfect solution, and this analysis could be strengthened in future work by a more detailed examination of prosodic features of the fillers in this language and the extent to which placeholder fillers recycle their targets.

3.3 Pro-verb uses of *mashiti*

Hayashi & Yoon (2006: 500) observe that (demonstrative) placeholder fillers are necessarily referential in that they refer to a “yet-to-be-specified lexical item” and are frequently replaced by the targeted lexical item later in discourse. This means that placeholder fillers are forward-looking.¹² Some uses of *mashiti*, however, are

¹²Hayashi & Yoon (2006: 500) note that while the forward looking nature of demonstrative place-holders makes them functionally similar to cataphora, the motivation is different. Placeholder fillers are motivated by hesitation and lexical retrieval, while cataphora is a deliberate discourse planning device.

not forward-looking, but instead look “backwards” at a previously uttered target. These uses of *mashti* are integrated into the syntax of the surrounding utterance – as evidenced by the presence of verbal morphology – and serve anaphorically to reiterate a previously uttered action or event. Thus, I label this function of *mashti* as a “resumptive pro-verb”. Pro-forms are semantically light forms that act as substitutes for more semantically substantive and discourse accessible elements (Schachter & Shopen 2007: 31). Thus, as a pro-verb, *mashti* reiterates or summarizes a previously uttered predicate. Table 6 shows the anaphoric targets of the pro-verb uses of *mashti* in the corpus. The majority target finite verbs and converbs.

Table 6: Target types of placeholder *mashti* ($N=16$)

<i>Target type</i>	<i>Count</i>
deverbal noun	1
clause	3
converb	5
finite verb	7

A pro-verb *mashti* targeting a converb is shown in (13), wherein the use of *imashtisha* in (13b) refers to the previously uttered adverbial clause headed by *nanachisha* ‘hurting’ in (13a). Note that *imashtisha* also mirrors *nanachisha* by having the same-subject *-sha* suffix.

- (13) a. *o sino=gā kāy išpa puru-ta nana-chi-sha*
 or otherwise=TOP DIST.DEM urine container-OBJ hurt-CAUS-ss
 ‘or rather, when the bladder is hurting’
- b. *kanser ni-k a-naun chi chasna imashti-sha*
 cancer say-SBJ.NMLZ COP-3PL DIST.DEM like.that PROV-ss
 wañu-shka-y
 die-PERF.NMLZ-LOC
 ‘they call it “cancer” when, happening[hurting] like that, [one] is
 dying’ [qvz015:82]

The pro-verb *mashti* can still point to a previous event without fully mirroring the verb. Consider the use of *mashti* in (14b) which does not mirror a specific lexical item in (14a). The *mashti* in (14b) is suffixed by the perfective *-shka*, but

there is no *-shka* marked material in (14a). The pro-verb *mashti* is referring to the overall predication of (14a) as a completed action.

- (14) a. *washa-manda bultiya-ri-sha kuti pay-ta randi*
 after-ABL turn.around-REFL-SS again 3-OBJ COORD.CON
yaku-ma aysa-n @@@@
 river-ALL pull-3SG [laughter]
 ‘after that, he’s turning around, but again he pulls him to the river,
 hahahaha’
- b. *chasna mashti-shka-Ø washa*
 like.that PROV-PERF-3SG after
 ‘and after [he] had done [it] like that’ [qvz010:47–48]

Both instances of the pro-verb *mashti* in (13) and (14) are modified by the manner adverb *chasna* ‘like that’. Half of the 16 pro-verb *mashti* tokens are modified by *chasna* or another manner adverb *kasna* ‘like this’. In such constructions, pro-verb uses of *mashti* resemble manner demonstrative verbs, which – among other functions – make anaphoric reference to previously uttered discourse units while describing the way in which an event occurred (Guérin 2015). This is congruent with the use of *imashti* as a pro-verb with manner adverbs as demonstrated in (13) and (14). However, without the adverbs, pro-verb *mashti* simply makes reference to a previous event without specifying manner. Additionally, manner demonstrative verbs typically derive from demonstratives (Guérin 2015: 146), whereas pro-verb *mashti* derives from a noun phrase (cf. §5).

The discourse distance between the utterance and the referring pro-verb *mashti* can be quite large, as shown in (15). In (15a-b), the speaker recounts how she became friends with a female forest spirit and her husband. She is then about to continue the story in (15e) when her interlocutor (TDS) interrupts and asks a question for clarification, which prompts a response in (15d-f) (transcription and interlinear gloss lines are omitted for brevity). Then in (15g), the speaker uses *mashtishka* to refer to the events of the utterance in (15a-b) and thus gets the narrative “back on track”.

- (15) a. *kunguri-sha kipiri-naku-sha ña*
 kneel-ss hug-RCPR-SS then
 ‘kneeling, and hugging each other, then’

- b. *amiga* *tuku-ra-nchi* *chi* *pay kari-wan=bas*
 female.friend become-PST-1PL DIST.DEM 3 husband-COM=CONJ
 ‘we became friends, and with her husband too’
- c. *chi* *washa ni-ra-*Ø
 DIST.DEM after SAY-PST-3SG
 ‘after that she said:’
 TDS: with her husband?
- d. ‘yes, with her husband’
- e. ‘the [spirit] woman and her husband’
- f. ‘her hair was down to here, the same as [my hair] down to here, she had long, full hair’
- g. *chi-* *chi* *mashti-shka-*Ø *washa*
 DIST.DEM[false start] DIST.DEM PROV-PERF-3SG after
 ‘that-, after that had happened’
- h. *chi-manda* *apa-wa-ra-*Ø
 DIST.DEM-ABL take-1.OBJ-PST-3SG
 ‘she took me’ [tds_qvz045:149–155]

The anaphoric nature of pro-verb *mashti* means that it cannot be a placeholder filler used in contexts of hesitation and disfluency, if the primary functional criterion of such fillers is the hesitating/delaying function in the service of lexical recall. In the pro-verb uses of *mashti* the referential target is uttered before the *mashti* is and, thus, is not likely to signal hesitation in recalling a lexical item since the target was already produced and is thus more immediately accessible in the speaker’s memory.

The differences between a placeholder filler *mashti* and a pro-verb *mashti* is thus whether or not it occurs in an episode of disfluency, and the direction of reference. This can be sometimes difficult to tease apart. Consider (16), wherein a speaker was shown a cartoon animation and is asked to recount the events shown therein.¹³ The *mashtira*, as a finite verb in (16b) could be a pro-verb referring back to the predicate, such as that of (16a): ‘the catfish followed the mouse’. Alternatively, *mashtira* could be an independent placeholder, or a placeholder

¹³The cartoon animation in question was a public domain Tom and Jerry short (Barbera & Hanna 1947).

filler that buys time for the lexical retrieval of *makara* ‘hit’ in (16c). In its morphology, *mashtira* could be mirroring either the preceding *katinakura* ‘[it] followed’ in (16a) or the following *makara* ‘[it] hit’ in (16c).

- (16) a. *washa=lla kati-naku-ra-*Ø (pause—632ms)
after=LIM follow-RCPR-PST-3SG

chi atun bagri
DIST.DEM big catfish
‘just after that, [he] followed [the mouse]... that big catfish’
- b. (pause—1691ms) *chi misi a-u-shka-*Ø *mayan-bi*
DIST.DEM cat COP-PROG-PERF-3SG close-LOC

(pause—699ms) ***mashti-ra-*Ø**
PH/PROV-PST-3SG
‘... close to where that cat was... [the catfish] was there/did it’
- c. (pause—2203ms) *chi bagri chupa-wan* *maka-ra-*Ø
DIST.DEM catfish tail-INST hit-PST-3SG
‘... that catfish hit [the cat] with its tail’ [qvz022:87–89]

I presented (16) as text and some contextual background information to a consultant and asked him what he thought *mashtira* referred to. With only the text to work with, the consultant interpreted (16b) as ‘[the catfish] was there close to where the cat was’, and specifically interpreted *mashtira* as ‘[the catfish] was there’, and thus an independent placeholder. To another consultant, I presented an audio clip of (16) with some contextual background information. She – by contrast – interpreted *mashtira* as referring to *makara* in (16c), and thus, as a placeholder filler.

Crucial to the difference in interpretations is the fact that one consultant had only the text to work with, while another had the audio clip. Note that in (16) there are several pauses, some of which are around two seconds in length. Because the speaker was recounting the events of the cartoon animation she was just shown from memory, she likely experienced some difficulty in recalling exactly what occurred in the cartoon animation; hence the disfluency events. Thus, the second consultant who had access to the audio clip of (16), noticed the pauses which likely contributed to her interpretation of *mashtira* as a placeholder. The first consultant, with only the textual representation of (16) to work with (without indications of disfluency), assumed a non-hesitative use of *mashtira*.

As is shown by (16) and the consultants' interpretation of it, morphosyntactic context may not always be sufficient to distinguish between placeholder and pro-verb uses of *mashti*. Pauses, as a correlate of disfluency, can be a useful factor. In §4 I expand further on this idea and present other correlates of disfluency that may assist in distinguishing between the different functions of *mashti*.

3.4 The form of *mashti*

Of the 157 tokens of *mashti* in the corpus, 75% take the form of *mashti* (N=117) and the remaining 25% as *imashti* with an initial /i/ vowel (N=40). I have not determined what, if any, difference between the presence or lack of the initial /i/ vowel makes with regards to the function or distribution of *mashti*. The examples in (17) and (18) come from the same speaker and same recording. In (17) *imashti* is used a hesitative filler, in (18b) *mashti* is also used as a hesitative filler.

- (17) *chi ni-k a-nau-ra imashti amu-yuk*
DIST.DEM say-SBJ.NMLZ COP-3PL-PST HES owner-PROP
'they call it, um, a (spirit) owner' [tds_qvz045:16]

- (18) a. *chi punzhana-wa-ta api-sha=lla ni-ra-ni ñuka*
DIST.DEM agouti-DIM-OBJ catch-1.FUT-LIM say-PST-1SG 1
‘I will just catch that little agouti’ I said’
- b. ***mashti***
HES
'um'
- c. *kay*
DIST.DEM
'here'
- d. *Montalvo-ma apa-mu-kpi*
Montalvo-ALL take-TRL-C-DS
'bringing it to Montalvo' [tds_qvz045:40–43]

The example given previously in (9) (cf. §3.2), shows *mashti* and *imashti* used as placeholders in the same utterance, which also comes from the same speaker and recording as (13) and (14). The same speaker also uses both forms for proverbs in the same recordings. Compare (15) in §3.3, which uses *mashti* as a pro-verb with the nominalized pro-verb *imashti* in (19), which comes from the same recording.

- (19) *chi-y kunguri-ra-nchi kasna*
 DIST.DEM-LOC kneel-PST-1PL like.this

imashti-shka-y

PROV-PERF.NMLZ-LOC

‘we kneeled there, and when we were doing it [kneeling] like this’
 [tds_qvz045:148]

All of these examples come from the same speaker (BD), who uses *mashti* and *imashti* interchangeably. In fact, BD accounts for 38 of the 40 *imashti* tokens, the other two come from LC, who also uses both forms. While I have not yet accounted for the variation between the two forms, the *imashti* form with the initial /i/ likely represents the “older” or “original” variant which I discuss in §5.

4 Disfluency correlates of *mashti*

As shown in §3, I generally consider uninflected instances of *mashti* to be hesitative fillers, while inflected instances of *masthi* can be categorized into placeholder filler (including use as an independent placeholder), and pro-verb, based on the presence or absence of a referent and the direction in discourse to that referent (before or after the utterance of *mashti*). However, these criteria are rough guidelines, as evidenced by several examples in §3. In fact, the identification of a given use of *mashti* in ambiguous morphosyntactic circumstances hinges on whether or not the speaker is experiencing a disfluency episode. This raises the following question: how do we know that the speaker is experiencing disfluency? This section explores the correlates of disfluency in the vocal-aural and visuo-spatial modalities and how those correlates interact with the use of *mashti*.

For an initial illustration, consider the following examples in Figure 4 and (20), and Figure 5 and (21), which come from a recording session wherein a speaker is interpreting the designs in her and her sister’s ceramic art for two researchers. These examples show uninflected instances of *mashti* that are ambiguous in terms of whether they are hesitative or placeholder fillers. However, the two examples are different in terms of the associated disfluency phenomena that co-occur with the use of the *mashti* fillers.

In still (i) of Figure 4 the speaker is holding a painted ceramic bowl that represents a type of plant and responds to questions from researchers about what the different painted patterns and shapes represent. In still (ii) of Figure 4, she points to a painted flower in the center of the bowl but experiences a disfluency while trying to articulate that the arrangement of shapes represents a flower. In

(20a), she produces a false start and *mashти* during her disfluency episode. Then in (20b) she resolves the disfluency and identifies the shape as the plant's flower. The use of *mashти* in (20a) can be interpreted as either a hesitant filler, or as a placeholder targeting the uninflected noun *sisa* 'flower' in (20b).



Figure 4: A speaker experiences disfluency while describing ceramic art (qvz017, 1:32–1:37), photo © 2019 Alexander Rice.

(20) a.

(ii)

kay- *kay* *mashti* *a-k=ka*
 PROX.DEM[false.start] PROX.DEM HES/PH COP-SBJ.NMLZ=TOP

a-u-n
 COP-PROG-3SG
 'this- this um/whatchamacallit is being a...'

b. *****

pay-ba *sisa*
 PROX.DEM=TOP 3-GEN flower
 'this is its flower' [qvz017:23–24]

The example in Figure 5 and (21) is similar to the example of Figure 4 and (21). The speaker, with a different ceramic piece, is likewise pointing to one of the designs painted inside of the bowl and experiences a disfluency while explaining to the researcher what it represents. In this disfluency episode, the speaker uses an uninflected *imashti* the final vowel of which is lengthened for 130ms. The use of *imashti* here may be that of a hesitative filler, or a placeholder filler targeting the following uninflected noun *apangura kiru* ‘crab claw’.



Figure 5: A speaker experiences disfluency while pointing to designs in ceramic art (qvz017, 1:58–2:01), photo © 2019 Alexander Rice.

- (21) *chi imashtii apangura kiru m=a-u-n*
 DIST.DEM HES/PH(130ms) crab claw EV.S=COP-PROG-3SG

kay=ga

PROX.DEM=TOP

‘that um.../whatchamacallit, is a crab claw, this one’ [qvz017:29]

As is shown in §4.3, some disfluency phenomenon like prosodic lengthening may be more associated with hesitant fillers compared to placeholder fillers and pro-verbs. Thus, the use of *mashti* in (20) – which is used without prosodic lengthening – may more likely be a placeholder filler compared to the *mashti* in (21).

This section goes over other correlates of disfluency in the interest of evaluating whether they can be used to better identify ambiguous uses of *mashti* with regards to its function as filler or pro-verb, and hesitant versus placeholder. Crucially, this analysis is multimodal in that I consider disfluency phenomena articulated not only by the vocal tract but also as articulated by other parts of the body. §4.1 goes over the disfluency correlates of *mashti* in the vocal-aural modality. §4.2 examines the visuo-spatial disfluency correlates of *mashti*. In §4.3, I look at how frequently each of the disfluency correlates occur with the different functions of *mashti* and offer some discussion on the patterns observed therein.

4.1 Disfluency correlates in the vocal-aural modality

The following disfluency correlates occur in the vocal-aural modality, which is to say is that the following disfluency phenomena are produced by the vocal tract: false starts (§4.1.1), pauses (§4.1.2), and prosodic lengthening (§4.1.3).

4.1.1 False starts

Fox Tree (1995: 710) identifies false starts as a disfluency phenomenon that occurs “when speakers start to say something, but then decide to abort their utterances and begin again. Fox Tree (1995) refers to the aborted articulation as the “false start”, and the correctly articulated lexical item as the “fresh start”. In (22), the speaker begins to utter the word *pasu* ‘mountain avocado’¹⁴ and articulates the first syllable /pa/, which is then aborted. This is the false start. The word is fully articulated later as *pasu*, and, thus, represents the fresh start.

- (22) *pa-* *mash**ti*** *pasu* *muyu-ta ña*
[false start] HES mountain.avocado fruit-OBJ DSC.PRT
‘the pa-, um, the *pasu* fruit’ [jbn_qvz002:124]

¹⁴The fruit of the tropical tree *Gustavia macarenensis*.

An uninflected *mashti* follows the aborted articulation (false start) of *pasu*. It is unlikely that this *mashti* is a placeholder filler targeting *muyu*, because the latter is inflected as an object, but it is possible that this *mashti* is a placeholder targeting the noun *pasu*, which acts as a modifier to *muyu* ‘fruit’. Navarretta (2016: 59) observes that hesitative fillers in Danish can occur in the same disfluency episode as false starts. Thus, the lack of clear syntactic integration plus the presence of a false start would make a better case for classifying this instance of *mashti* as a hesitative filler instead of a placeholder filler.

Mashti itself can also be subject to a false start. In (23) the articulation of /ma/ is likely a false start of the hesitative *mashti* that is correctly articulated afterwards.

- (23) *ma-* *mashti wañu-shka mikya*
 [false start] HES die-PERF.NMLZ auntie
 ‘uh- um, [my] deceased auntie’ [jbn_qvz002:263]

Every instance of *mashti* in the corpus was annotated for the presence or lack of a false start in the same intonational unit. The total counts are given in Table 7.

Table 7: Occurrence of false starts with *mashti* in corpus

<i>mashti</i> function	Occurs with false start	Occurs without false start
hesitative filler	10	109
placeholder filler	0	22
pro-verb	2	14

Overall, there are only 12 instances of false starts occurring in the proximity of *mashti*. Most occurred with what I deemed to be hesitative fillers, while two occurred with pro-verbs.

4.1.2 Pauses

A silent pause – in which the speaker stops talking – is the “simplest way to hesitate” (Lickley 2015: 456). However, as silence is a feature of fluent speech, it can be difficult to define how much silence constitutes a pause. Silence between intonational units in fluent speech may be longer than silence between words. Lickley (2015: 456–457) gives an overview of the problems in distinguishing between “disfluent” pauses and “normal” pauses and shows that annotators make do with subjective perceptual judgement.¹⁵

¹⁵Lickley (2015: 457) also points to making use of multiple annotators and estimating inter-rater reliability as the “safest approach” for identifying disfluent pauses. I did not take this approach

I likewise identified disfluent pauses based on my own subjective and perceptual judgement.¹⁶ The pauses I identified were between 500–2000ms in length. My identification of pauses also relied on the presence of other disfluency factors and overt signs of hesitation, which I discuss elsewhere in §4.2 (excessive blinking, gaze aversions, gesture freezes, etc.). I marked each instance of *mashti* that occurred with a pause either in or adjacent to the *mashti* bearing intonational unit.¹⁷ An example is given in (24b) where a pro-verb *mashti* is followed by a 2.5 second pause.

- (24) a. *alku kishpi-sha ni-sha taula-guna-y api-ri-ra-*∅
dog escape-ss say-ss plank-PL-LOC grab-MID-PST-3SG
'the dog, wanting to escape, grabs hold of the planks (at the dock)'
- b. *chi aysa-sha mashti-kpi* (pause—2530ms)
DIST.DEM pull-ss PROV-DS
'and so pulling, (he's) doing that while....'
- c. *chi ukucha*
DIST.DEM mouse
'that mouse'
- d. *ña mayan-bi a-u-shka-y pay-ta*
DSC.PRT close-LOC COP-PROG-PERF.NMLZ 3-OBJ
shuk kaspi-ta ku-gri-ra-∅
one stick-OBJ give-TRLC-PST-3SG
'being close by, went to give [the dog] a stick' [qvz022:121–123]

Table 8 shows how each of the *mashti* functions occurs with a pause. Overall, there are more instances of pauses occurring with *mashti* in the corpus than false starts, and hesitant uses of *mashti* would appear to occur more with pauses

for this research, due to the limited availability of Northern Pastaza Kichwa speakers and fieldwork funding constraints. Nevertheless, I acknowledge that future work in this vein would benefit greatly from including Northern Pastaza Kichwa speakers in annotating disfluency, and their ratings might differ from my own.

¹⁶I am not a native speaker of Northern Pastaza Kichwa, but I do possess a moderate fluency in the language from having taken formal language classes in Ecuadorian Quechua and from working with Northern Pastaza Kichwa speakers for over a decade. I believe my ability to speak the language is sufficient enough to at least identify disfluent pauses.

¹⁷See Ponsonnet (2025 [this volume]) and Rose (2025 [this volume]) for the interaction between pauses and fillers/placeholders in Dalabon and Teko, respectively.

than placeholders or pro-verbs do. This would point to the presence of pauses as a better indicator of disfluency compared to false starts. Although, as discussed in §4.3, the distribution of the data severely limits quantitative assessments.

Table 8: Occurrence of pauses with *mashti* in the corpus

<i>mashti</i> function	Occurs with pause	Occurs without pause
hesitative filler	24	95
placeholder filler	5	17
pro-verb	1	15

4.1.3 Prosodic lengthening

Prosodic lengthening is the non-phonemic prolongation of a syllable which can often occur in episodes of disfluency (Lickley 2015: 458). In (25b), the final /i/ vowel of a hesitative filler *mashti* is held for 418ms.

- (25) a. *chasma a-sha=mi pay-guna libru-ta*
 like.that COP-SS=EV.S 3-PL book-OBJ
 tiya-k-ta tupa-nau-ra
 be-SBJ.NMLZ-OBJ find-3PL-PST
 ‘being like that, they found the existing book’
- b. ***mashthii*** (pause—7625ms)
 HES(418ms)
 “um...”
- c. *dos mil ocho*
 two thousand eight
 ‘[in] 2008’ [qvz026:133–135]

As is the case with silent pauses in disfluent speech versus fluent speech, determining whether the duration of a syllable in disfluent speech is actually “prolonged” is methodologically fraught (Lickley 2015: 458). As with the silent pauses, I used my own subjective and perceptual judgement to determine whether or not a given instance of *mashti* was prosodically lengthened. However, I corroborated this approach by comparing the durations of the final vowel in each *mashti* token

with a sample of durations of the same vowel in a similar phonetic environment but in a non-hesitative context. This was especially necessary given that placeholder and pro-verb uses of *mashti* can take a variety of suffixes and enclitics. The durations of non-prolonged and prolonged final /i/ vowels may be different than that of final /a/ vowels. Some suffixes attached to placeholder and pro-verb uses of *mashti* also possess coda consonants and thus need to be compared to other phonetically similar material with the same coda consonant.

There are 18 different phonemic forms that represent the word final syllables of all the *mashti* tokens in the corpus. To define the duration for non-prolonged vowels and codas of these syllables I semi-randomly selected ten words for each syllable type in contexts that I perceived to be fluent speech and measured the duration of the final vowel and coda consonant (if present). I then used the average of the ten values as the baseline non-prolonged duration. These baseline values are given in Table 9.¹⁸

I then compared the duration of the final syllable of each *mashti* token with the corresponding value in Table 9. If the duration was greater than the baseline value, I marked it as prolonged. To illustrate, the sampled average duration for the /a/ vowel in a word final /wa/ syllable is 78.4ms. In (26) the placeholder filler *imashti* carries the same-subject purpose clause suffix *-ngawa*, the final syllable of which is /wa/. The duration of the /a/ vowel is 119ms, which is longer than the 78.4ms baseline, and is thus marked as “prolonged”.

- (26) a. *mana chasna imashti-*ngawaaa**
NEG.INTJ like.that PH-SS.PURP(119ms)
'no, in order to whatchamacalllit like that'
- b. *ali- ali-chi-*ngawa mana pay sagra-shka-*Ø*
[false.start] good-CAUS-SS.PURP NEG 3 curse-PERF-3SG
'fix- in order to fix the curse that was done' [qvz015:17-18]

Table 10 shows how many tokens of each *mashti* function were prosodically lengthened. As is the case with silent pauses, the hesitative filler use of *imashti* would appear to attract prosodic lengthening more so than the placeholder filler or pro-verb uses of *mashti*. In fact, the majority of hesitative *mashti* tokens were prosodically lengthened.

¹⁸The reference words I used and the individual durations can be found in the “ref_durations.xlsx” document in the OSF drive: <https://osf.io/g84mv/>. The reference words were selected semi-randomly. In general I opted for words that occurred at the end of intonational units, which is the position that many but not all *mashti* tokens occur in. I also tried to vary speakers as much as possible.

Table 9: Reference durations for word final phonemic sequences in fluent speech

Word final syllable	Average duration of final vowel and coda (ms)
/tʃun/	98.3
/da/	97.5
/ga/	111
/ka/	104.8
/kpi/	120.8
/na/	87.9
/ni/	80.9
/ra/	76.3
/ʃa/	123.5
/fi/	108.5
/ʃkai/	148.2
/ta/	94.9
/tas/	166.4
/ti/	73.7
/wa/	78.4
/wan/	144.3
/was/	211.2
/wi/	91.9

Table 10: Occurrence of prosodic lengthening in *mashti* tokens in the corpus

<i>mashti</i> function	With prosodic lengthening	Without prosodic lengthening
hesitative filler	98	21
placeholder filler	13	9
pro-verb	5	11

4.2 Visuo-spatial modality correlates of disfluency

The following disfluency correlates occur in the visuo-spatial modality, which involves the use of bodily articulators outside of the vocal tract: the use of gaze (§4.2.1), instances of excessive blinking (§4.2.2), the use of manual gestures (§4.2.3), and gestural disfluency (§4.2.4).

4.2.1 Gaze aversion

Gaze and head movement are an active component of speech and serve a variety of functions such as turn management, attention signalling, attention-checking, and viewpoint shifting (Rossano 2012; Sweetser & Stec 2016; Thompson & Suzuki 2014). With regards to disfluency events, a shift in gaze away from the addressee (gaze aversion) can co-occur with a word search (Goodwin & Goodwin 1986) and can invite/disinvite assistance from the interlocutor (Hayashi 2003). In particular, prior research has shown that gaze aversions often co-occur with fillers as a means of “holding the floor” during disfluency episodes (Brône et al. 2017; Feyaerts et al. 2017).

Speakers of Northern Pastaza Kichwa likewise use gaze aversions and fillers during disfluency events, as is shown in Figure 6 and (27). In (27a) and (i) of Figure 6, the speaker’s gaze is directed to the left (her right) whilst narrating. She then experiences a disfluency when attempting to introduce a new character to the narrative. The disfluency is initiated by using the discourse particle *ña* as a hesitative filler and a pause in (27a). At the same time, she abruptly moves her gaze to the right (her left) and holds it there while hesitating, as shown in (ii) of Figure 6. The use of the repeated “==” symbols in (27) shows the approximate duration of the speaker’s gaze relative to discourse, and the carets show the approximate direction relative to the camera. The averted gaze co-occurs with the use of the hesitative filler *mashti* in (27b). In (27c) and (iii) of Figure 6, the speaker recalls the name of the character she wants to introduce and moves her gaze back to the left (her right) to her interlocutor.

- (27) a. (i) (ii)
=====>
chi-manda=mi *ña* (pause—820ms)
DIST.DEM-ABL=EV.S DSC.PRT
'and from there, um...'

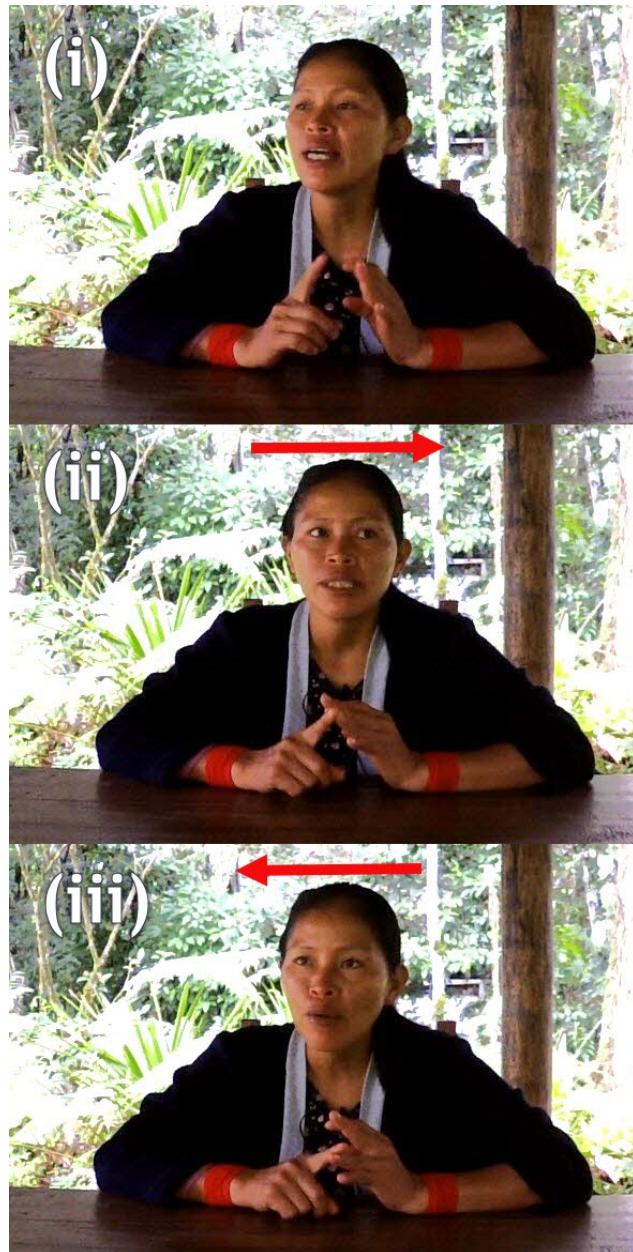


Figure 6: A speaker averts her gaze while hesitating (qvz026, 5:06–5:11),
photo © 2019 Alexander Rice

b. ===>

mashti

HES

'um'

c. (iii)

<=====

Andrea Dagua

Andrea Dagua

'Andrea Dagua' [qvz026:101–103]

I annotated gaze aversion as an abrupt movement of the eyes and head away from one direction towards another. I did not measure the duration of each averted gaze, only whether or not it occurred, and I only annotated those that occurred in the presence of *mashti*, either in the same intonational unit, or in an adjacent intonational unit (before or after) in the same disfluency event. Table 11 shows how many *mashti* tokens of each type co-occurred with a gaze aversion. Across all types, gaze aversion. Overall, each type of *mashti* occurred more with gaze aversion than without.

Table 11: Occurrence of gaze aversions with *mashti* tokens

<i>mashti</i> function	With gaze aversion	Without gaze aversion
hesitative filler	81	37
placeholder filler	16	6
pro-verb	10	6

4.2.2 Excessive blinking

Research in neuroscience has shown that blinking can serve communicative functions in speech (Mandel et al. 2014, Nakano & Kitazawa 2010). With regards to the interaction between disfluency and blinking, Lickley (2017: 382) briefly mentions that excessive blinking can be associated with stuttering. Hömke et al. (2017) show that addressees, when faced with speakers experiencing disfluency, can use long blinks to signal to the speaker that repair is not needed and additionally signal a disinterest in "taking the floor". Beyond these studies, the interaction between disfluency and the blinking of the speaker has not yet been explored in detail as far I am aware.

In the present corpus, disfluency episodes are sometimes accompanied by multiple blinks performed in rapid succession. I annotated instances of *mashti* as being associated with excessive blinking when two or more subsequent blinks were performed in vicinity of the articulation of *mashti*, either in the same or adjacent intonational unit in the same disfluency event. In (24), the speaker blinks four times while uttering *mashti*. The “^” symbol above the transcription line indicates one blink. Repeated blinking is somewhat difficult to show in a succinct way with static images so, Figure 7 shows only one of the four blinking sequences performed concurrently with the uttering of *mashti* in (28).



Figure 7: A speaker blinks several times in quick succession while uttering *mashti* (lwc_qvz002, 0:59–1:03), photo © 2016 Lisa Warren Carney.

(28)

			^^^^
chi-ta	riku-sha	mashtii	chi-ta
DIST.DEM-OBJ	see-ss	HES(395ms)	DIST.DEM-OBJ
			dream-SBJ.NMLZ
<i>likcha-ri-ra-ni</i>			
wake.up-REFL-PST-1SG			
'seeing that, um, dreaming (about) that, I woke up'	[lwc_qvz002:13]		

Table 12 shows how many *mashti* tokens occurred in the presence of an excessive blinking episode. Overall, excessive blinking does appear to occur as frequently in disfluency events involving *mashti* as gaze aversions and prosodic lengthening.

Table 12: Occurrence of excessive blinking with *mashiti* in corpus

<i>mashiti</i> function	With excessive blinking	Without excessive blinking
hesitative filler	29	87
placeholder filler	1	22
pro-verb	2	14

4.2.3 Manual gestures

Research in psychology and neuroscience has shown that manual gestures often occur during disfluency events, and more specifically, some have argued that manual gestures may aid in lexical retrieval (Butterworth & Beattie 1978; Rauscher et al. 1996; Krauss & Hadar 1999; Pyers et al. 2021; Osorio et al. 2024). Research in this vein has come under the banner of the “Lexical Retrieval Hypothesis” which, generally, posits that manual gestures facilitate lexical retrieval. This would suggest that filler words like *mashiti* and manual gestures might co-occur more often than not.

The Lexical Retrieval Hypothesis is not a consensus position, however: other studies have presented challenges for some of its assumptions. For example, Christenfeld, Schachter, and Bilous 1991 show that the use of a gesture during a disfluency episode may preclude the use of a filler word. Other research has challenged the Lexical Retrieval Hypothesis more directly in finding that gestures are not especially prevalent in disfluent speech versus fluent speech (Graziano & Gullberg 2018; Hoetjes et al. 2014; Kisa et al. 2022).

In the interest of gaining additional insight into the relationship between gesture and disfluency, I marked each use of *mashiti* that contained a co-occurring manual gesture. One such example is given in Figure 8 and (29), wherein the speaker is describing how characters in a narrative made spears. She uses a gesture wherein her left arm is extended out in front her torso with the left palm oriented upwards. Her right hand is held in a loose cupping shape and she moves her right hand up and down over her left hand repeatedly. In the context of spear-making, this gesture is likely an iconic gesture wherein the speaker’s left hand represents the piece of wood being shaped into a spear, and the right hand depicts a hand holding a sharp instrument such as a knife or other scraping tool. The repeated movement of the right hand over the left thus represents a process of chamfering or scraping the piece of wood to create the shaft of the spear. The speaker begins to use the gesture in still (i) of Figure 8 and (29b) and continues to perform the gesture when she experiences a disfluency and pauses after uttering

sumakshi ‘reportedly nice/nicely’.¹⁹ During the pause in speech, she continues to perform the spear-making gesture.

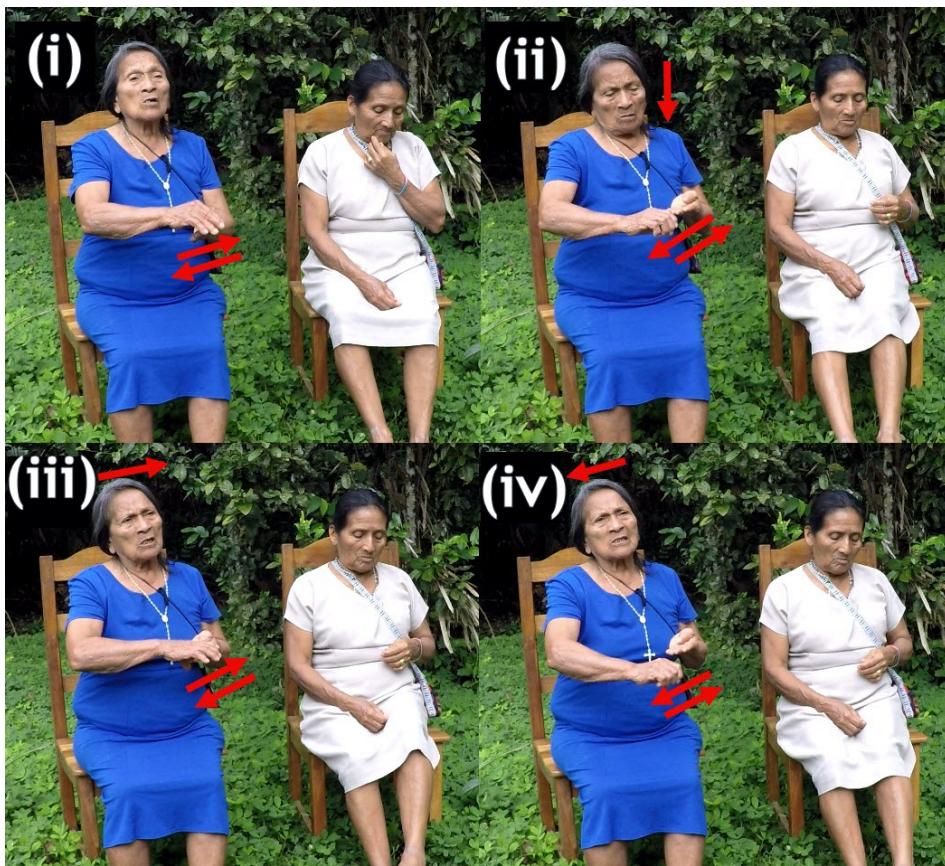


Figure 8: A speaker uses a manual gesture depicting the chamfering of a wooden spear shaft while hesitating in recalling the word for “spear” (jbn_qvz003, 2:27–2:37), photo © 2015 Alexander Rice.

- (29) a. *shina piti-sha*
 like.this cut-ss
 ‘cutting like this’

¹⁹ *Sumak* ‘nice/beautiful’ can modify both nouns and predicates, thus the aborted utterance may have been something like; “they had made reportedly nice spears” or “nicely they had made spears, reportedly”.

b. (i)

sumak=shi (pause—642ms)

nice=EV.O

‘nicely, reportedly...’

c. (ii)

v =====

pay-guna (pause—966ms)

3-PL

‘they...’

d. (iii) (iv)

====>

mashti lansa-ta ra-shka-una

HES spear-OBJ do-PERF-3PL

‘um, they had made spears’ [jbn_qvz003:31–35]

In (29c), she starts the narration again, continuing to perform the gesture, then pauses her speech and freezes her gesture. In (ii) of Figure 8, she averts her gaze from her interlocutor to look down at her hands. Then she returns her gaze to her interlocutor and starts the gesture again, while still staying silent. In (iii) of Figure 8 and (29d) she utters a hesitative filler *mashti* and averts her gaze to the right (her left) while continuing to perform the manual gesture. Finally, in (iv) of Figure 8 and (29d) she returns her gaze to her interlocutor and escapes the disfluency event by uttering *lansa rashkauna* ‘they had made spears’.

In Figure 8 and (29), it would appear that the speaker’s hands essentially “knew” what the target utterance was before the speaker’s vocal tract did.²⁰ The spear-making gesture starts approximately 4.5 seconds before the speaker resolves the disfluency and articulates the spear-making event. Unable to recall a word or experiencing difficulty in articulating a phrase, a speaker may thus appear to use their hands to help “capture” or “organize” the target material, as is argued by the Lexical Retrieval Hypothesis. Additionally, gesture use during a disfluency may serve as an additional turn holding device and communicate the target material to the interlocutor visually.

²⁰See Döhler (2025 [this volume]) for similar examples of gestures accompanying the place holder *bäne* in Komnzo.

Table 13 shows the *mashti* tokens that were uttered concurrently with a manual gesture. Overall, roughly half of the *mashti* tokens for each category occurred with a manual gesture.

Table 13: The occurrence of manual gestures with *mashti* in the corpus

<i>mashti</i> function	Uttered with gesture	Uttered without gesture
hesitative filler	75	43
placeholder filler	11	11
pro-verb	9	7

I should note, however, that I did not take into the account the types of gestures that occurred with *mashti*. Nor does this data weigh in on the validity of the Lexical Retrieval Hypothesis since the rate or types of gestures in fluent speech were not considered. Here I am only interested in whether one of the different functions of *mashti* co-occurs with manual gestures compared to the others. As is shown in Table 13, the presence or absence of a gesture is roughly equal between the different *mashti* types.

4.2.4 Gestural disfluency

Like speech, manual gestures are also subject to disfluency effects (Seyfeddinipur 2006). In an experimental setting, Betz et al. (2023) observed several of the following types of gestural disfluency, including cancellation (aborting a gesture), pauses (holds), slow-downs, and metaphoric gestures. In the corpus data, I found three types of manual gesture disfluency that occurred in proximity to a *mashti* token.

- Deactivation: The end of a manual gesture happens to coincide with a disfluency episode involving *mashti* and no gestures are used until the disfluency is resolved.
- Hold: A gesture is paused or “frozen” mid-stroke as a speaker experiences a disfluency episode. The gesture resumes when the disfluency is resolved.
- Cycle: A gesture in which the hand or finger traces a circular path continuously while the speaker experiences a disfluency episode. The speaker stops the gesture or switches to a new gesture when the disfluency is resolved.

An example of gestural deactivation is given in Figure 9 and (30). In (i) and (ii) of Figure 9, the speaker is actively gesturing while telling a story. In (30b) she experiences disfluency and stops gesturing as shown by her hand dropping to her side in (iii) of Figure 9. The gestural deactivation is concurrent with the use of the hesitative *mashti* and the pause in speech in (30b). Note also that during the pause, no gestures are used while the speaker hesitates. Thus, there is a “gestural pause” parallel to the prosodic pause. When the disfluency is resolved in (30c) the speaker begins gesturing again.

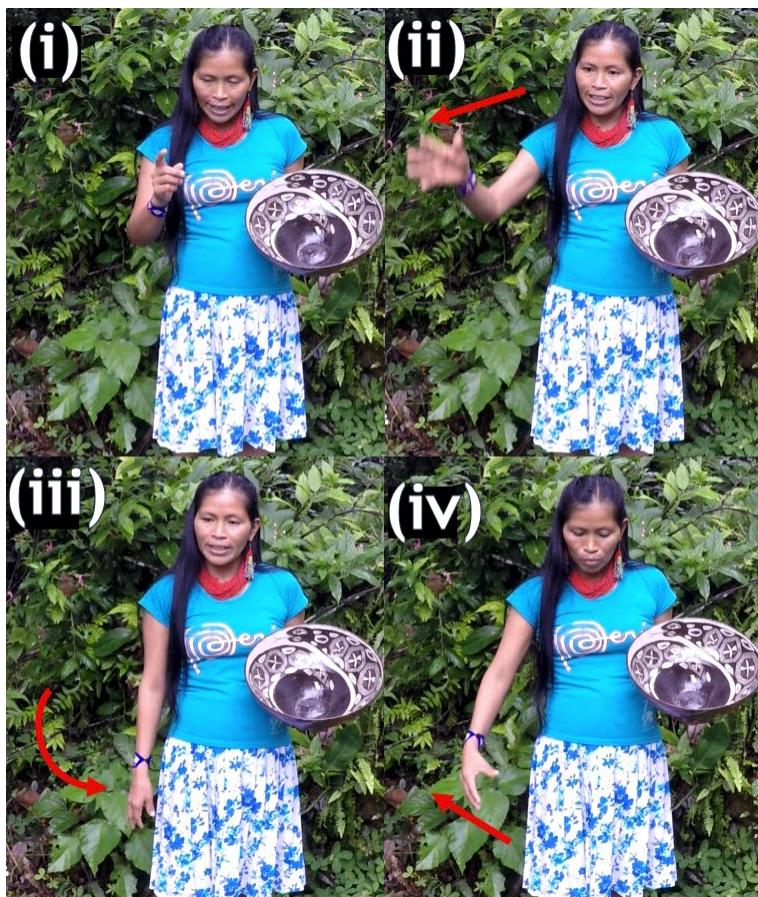


Figure 9: A speaker deactivates her gesture while experiencing a disfluency episode (tds_qvz034, 5:23–5:29), photo © 2015 Alexander Rice.

- (30) a. (i)
- ****
- chi kaya-ndi*
DIST.DEM tomorrow-ADD
'and then the next day'
- b. (ii) (iii)
- *****
- ri-ra-nchi mashti* (pause—710ms)
go-PST-1PL HES
'we went, um...'
- c. (iv)
- *****
- riku-kpi chi muriti-y*
see-DS DIST.DEM peach.palm.grove-LOC
'and there in that peach palm grove we saw:' [tds_qvz045:72–74]

Figure 10 and (31) show an example of a gesture hold associated with disfluency. The speaker tells how some characters in a narrative ran to a large tree and spreads his arms out to show the relative width of the trunk in (ii) of Figure 10. However, the speaker also experiences a disfluency in recalling the name of the tree in (31b). While experiencing the disfluency, he holds the spread-arms gesture in (ii) of Figure 10 for nearly three seconds while struggling to recall the name of the tree in (31b–31c). In (iii) of Figure 10 and (31d), the speaker raises his hands and head slightly, which is another indicator of disfluency (Ozkan et al. 2023). The speaker then resolves the disfluency and recalls the name of the tree in (31d) and lowers his arms in (iv) of Figure 10 to resume the spread-arm gesture indicating the width of the tree trunk.

- (31) a. (i)
- *****
- chasna ña chaulla-ta miku-sha ri-shka-ta*
like.that then ideo:clean-ADV eat-ss go-PERF.NMLZ-ADV
- *****
- apa-naku-sha ri-kpi=ga*
take-RCPR-SS go-DS=TOP
'then like that, having cleanly eaten (everything), they had gone,
chasing (them)'

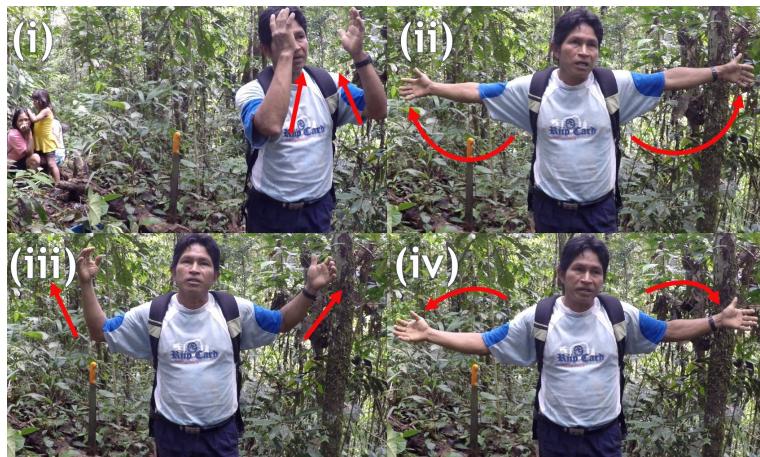


Figure 10: A speaker holds a gesture while experiencing a disfluency (tds_qvz031 9:46–10:50), photo © 2015 Alexander Rice.

b. (ii)

******(hold—2942ms)

atun mashti kay (pause—1019ms)
big HES/PH PROX.DEM
'(there is) a big, umm, this...'

c. *****

ima m=a-n kay
what EV.S=COP-3SG PROX.DEM
'what is this (one)?'

d. (iii) (iv)

***** * *****

ñuka-nchi ni-nchi wayra kaspi
1-1PL say-1PL wind tree
'we call it "the wind tree"' [tds_qvz031:197–200]

Ladewig (2011) shows that German speakers make use of repeated circular movement gestures (cycle gestures) with their fingers or hands in contexts of “word searches” (i.e. disfluency). Such cycle gestures would be considered as “metaphoric” gestures by Betz et al. (2023). The form of the gesture – the circular motion of a hand or finger – is thought to represent “the continuous nature of the searching activity” (Ladewig 2011: 8).

I take such circular gestures in the context of disfluency as additional indicators of disfluency in Northern Pastaza Kichwa discourse. Figure 11 and (32) show a speaker using a cyclic gesture concurrently with the hesitative *mashti*. The example comes from a historical narrative about the speaker's ethnic background as a member of the Andwa ethnic group.²¹ In (32), she describes the conditions of one of the smallpox epidemics that occurred at the turn of the twentieth century and, more specifically, the smallpox symptoms that the Andwa people experienced. The speaker explains that the people often had to go naked because the pustules associated with smallpox that covered the body made it painful and unpleasant to wear clothes. More specifically, the pustules would stick to the clothes, and upon changing clothes, the pustules would be painfully ripped off the skin.

The speaker describes this by pointing to the back of her hand in (i) of Figure 11 to use her own skin as an example while describing the situation with the clothes in (32a). However, she experiences a disfluency in (32b) which is characterized by the use of a hesitative *mashti* with prosodic prolongation and a silent pause. While uttering *mashti* and pausing, she circles her index finger on her hand twice while hesitating, which I take to be a cyclic gesture. In (32c) she resolves the disfluency and moves ahead with the planned utterance and gesture of tapping her skin with her index finger to show how the pustules would stick.

- (32) a. (i)

dinu=ga kay llachapa-ta chura-ri-kpi=ga llachapa=ma
 rather=TOP PROX.DEM clothes-OBJ put-REFL-DS=TOP clothes-ALL

kasna
 like.this
 'rather, when putting on their clothes (it would stick) to their clothes'
 b. (ii)
***** (2 repetitions)
mashthii (pause—743ms)
HES(480ms)
 'um...'

²¹The Andwa territory is located on the Bobonaza river near the Ecuadorian military base in Montalvo of the Pastaza province of Ecuador. Historically the people in this territory spoke a now dormant Zaparoan language (probably Andoa [ISO 639-3: anb], endonym: Katsakáti).



Figure 11: A speaker cycles her finger index while hesitating (qvz026, 2:48–2:58), photo © 2019 Alexander Rice.

c. (iii) (iv)

aglluña-sha=ga tyapi *tyapi=shi*
stick-ss=TOP IDEO:sticking IDEO:sticking=EV.O

api-ri-k *a-shka-*∅
grab-REFL-SBJ.NMLZ COP-PERF-3SG

‘sticking, it was reportedly like *tyapi!* *tyapi!* [the smallpox pustules stick to the clothes and are ripped off when clothes are taken off]’
[qvz026:55–57]

Table 14 shows the *mashti* tokens that were uttered while the speaker experienced a gestural disfluency. Note that the total sets of tokens are less than those of the other disfluency correlate counts. This is because for a gestural disfluency to occur with *mashti*, there had to be a gesture present in the first place. Thus, the tokens in Table 14 represent a subset of the ‘Uttered with gesture’ tokens in Table 13.

Table 14: Occurrence of *mashti* with gestural disfluency

<i>mashti</i> function	With gestural disfluency	Without gestural disfluency
hesitative filler	48	27
placeholder filler	3	8
pro-verb	0	9

4.3 Summary

Table 15 shows the frequency of all the considered multimodal indicators of disfluency with the different *mashti* functions expressed as percentages. Due to the low *N* values of the placeholder and pro-verb uses of *mashti* the differences in percentages between the groups may be spurious. Additionally, recall that I use the correlates of disfluency to categorize ambiguous instances of *mashti* (cf. §3 and §4). Thus, evaluating the co-occurrence between the type of *mashti* and the disfluency correlates is somewhat circular, because – for some *mashti* tokens – the category was not assigned independently of the correlates. Nevertheless, there are some patterns in the data that can serve as jumping-off points for forming hypotheses for future investigations with more data and the use of independent and categorical criteria.

Table 15: Frequency of disfluency correlates with *mashti* functions

<i>mashti</i> function	H (<i>N</i> = 119)	PH (<i>N</i> = 22)	PRO-V (<i>N</i> = 16)
False start	8%	0	12%
Pause	20%	22%	6%
Pros. lengthening	82%	59%	31%
Gaze aversion	68%	72%	62%
Excessive blinking	24%	4%	12%
Gesture	63%	50%	56%
Gestural disfluency	40%	13%	0

Looking across the columns, we can see that prosodic lengthening, gaze aversion, gesture, and gestural disfluency co-occur with the filler uses of *mashti* more frequently than false starts, pauses, and excessive blinking. Upon comparing the *mashti* categories (rows) it would appear that prosodic lengthening, excessive

blinking, and gestural disfluency occur more with the hesitative filler *mashti* compared with the placeholder and pro-verb uses of *mashti*.²² Keeping in mind the low *N* value of the pro-verb *mashti* tokens, gaze aversion and gesture would appear to occur more frequently in the pro-verb context than false starts, pauses, prosodic lengthening, excessive blinking, and gestural disfluency.

The differences between the frequencies shown in Table 15 cannot be statistically tested due to the low *N* values of the placeholder and pro-verb instances of *mashti*. Additionally, the nature of the data presents the challenge of multicollinearity, because the tokens are drawn from a small number of speakers and some of the potential variables depend on other variables (i.e. gestural disfluency requires that a gesture is performed in the first place).

With the caveat that the proportions of the multimodal disfluency correlates of *mashti* as shown in Table 15 are likely not statistically significant, the suggestion that prosodic lengthening may be part of what distinguishes between the filler and non-filler uses of *mashti* is not unwarranted. Rose (2025 [this volume]) shows that the (hesitative) filler *baʔe* in Teko is more lengthened than the generic noun *baʔe* ‘thing/non-human’.²³ Vallejos Yopán (2023) also finds that the filler use of *este* as a hesitative filler in Amazonian Spanish is more associated with prosodic lengthening compared to the use of *este* as a demonstrative.²⁴

The results presented in this section provide the basis for some hypotheses that can be tested in future work with a larger corpus with higher *N* values for placeholder and pro-verb uses of *mashti*. Is the presence or absence of prosodic lengthening predictive of the filler or pro-verb use of *mashti*? Is the same true of gestural disfluency or the other multimodal disfluency correlates? If the trends presented here were to hold for a larger dataset and could be tested, I would argue overall that the case of *mashti* in Northern Pastaza Kichwa shows that multimodality plays a bigger role in language function than is usually appreciated. Nuckolls (2020) shows that manual gestures are an essential component of defining ideophonic words in Northern Pastaza Kichwa. It may be the case that bodily indicators such as gaze aversion and gestural disfluency may be essential in distinguishing the different functions of *mashti*.

²²Note that a subset of the placeholder *mashti* tokens are independent placeholders and not necessarily associated with disfluency like placeholder fillers are (cf. §3.2). With more tokens, it would be instructive to compare independent placeholders to the other categories.

²³Tupi-Guaranian, French Guiana [ISO 639-3: eme].

²⁴However, Hennecke & Mihatsch (2022) find that using prosodic cues to disambiguate filler and placeholder functions in the French filler/placeholders *truc* and *machin* is problematic.

5 The origin and developmental pathways of *mashti*

In the previous sections, I have discussed the different linguistic functions of *mashti* in discourse (as a hesitative filler, placeholder filler, and pro-verb), as well as the multimodal disfluency phenomena that may play a role in distinguishing between the discourse functions of *mashti*. In this section I discuss the origin of *mashti* and present the potential pathways of how *mashti* became associated with the disparate functions it now has.

The use of *mashti* as a filler is attested in other varieties of the Colombian-Ecuadorian branch of Quechua. Grzech (2016a) attests of *mashti* as a hesitative and placeholder filler in Upper Napo Kichwa – the sister variety of Northern Pastaza Kichwa – and it is used in varieties of Ecuadorian Highland Kichwa, Lower Pastaza Quechua (Peru), and Inga (Colombia) (p.c. Simeon Floyd, 2022). The form *imasti* is also attested as a filler word in Santiago del Estero Quichua (Argentina) (Albarracín de Alderetes 2009), which suggests that *mashti* has been used as filler word for centuries, at least as far back as pre-Spanish contact times in the fifteenth century.²⁵ ²⁶ The oldest attested source of *mashti* in any Quechuan variety that I am aware of comes from a nineteenth century dictionary of a variety of Ecuadorian Quechua spoken in the Azuay province of Ecuador (Cordero 1892),²⁷ wherein *imasti* and other, similar forms are listed with definitions that suggest filler functions. A scan of the page containing these entries is given in Figure 12.²⁸

The English translations for the relevant entries in Cordero's dictionary are as follows (English translations are my own):

²⁵The “Quechua II” branch includes northern and southern most extensions of Quechuan languages (in Colombia, Ecuador, Northern Peru, Southern Peru, Bolivia, Chile and Argentina), which correspond to the outermost regions of the former Inca Empire. The languages of the Quechua I branch inhabit the central and southern parts of Peru. The development and spread of the Quechua II varieties is thought to have occurred with the last major expansion of the Inca empire from Cusco in the early fifteenth century roughly a generation before Spanish colonizers arrived. The spread of Quechua into Southern Bolivia (and later into Argentina) is thought to have been approximately simultaneous with the arrival of Quechua in Ecuador. Thus, if *mashti* is used in Quechuan varieties as far south as Argentina and as far north as Colombia, we can surmise that *mashti* was used as filler as least as far back as this split in the fifteenth century. It does not appear that *mashti* is used in the Quechua I varieties (p.c. Simeon Floyd, 2022)

²⁶Consult Adelaar & Muysken (2004: 180–191) for an overview of the spread and development of the Quechuan language family.

²⁷This Quechua variety would correspond to the present day Cañari Kichwa, a variety of Ecuadorian Highland Kichwa.

²⁸Credit goes to Simeon Floyd for teasing out the geographic distribution of *mashti* as well as pointing out Cordero's dictionary to me.

- IMASTI**, n. Este, esta, esto, cuyo nombre no recuerdo.
- IMASHINA**, adv. Como; así como; de la manera que.
- IMASHINAMI**, adv. Lo mismo que *Imashina*.
- IMASHINAPISH**, adv. Como quiera; de cualquier modo.
- IMASHUTI**, n. Aquel; aquella o aquello cuyo nombre se me escapa de la memoria; pero lo indico vagamente.
- IMASHUTINA**, v. a. Hacer aquello cuyo nombre se me olvida en este instante; pero deseo hacer comprender de algún modo.

Figure 12: Excerpt from Cordero's dictionary attesting the use of *imasti* in Ecuadorian Highland Kichwa in the nineteenth century (Cordero 1892: 40).

- *imasti*, n. "This, that, this one, whose name I cannot recall."
- *imashuti*, n. "That; that one whose name escapes my memory; but which I vaguely indicate."
- *imashutina*, v. "Doing something which I do not remember the name of in this instant; but I would like to make intelligible in some way."

Some initial observations about the origin of *mashti* can be drawn from these dictionary entries. The definition of the form *imashuti* is very close to that of *imasti*. The form *imashuti* is transparently composed of the interrogative pronoun *ima* 'what' and *shuti* 'name'. In the present Northern Pastaza Kichwa corpus, *ima* and *shuti* occur together only once and are used to ask someone's name, as shown in (33). Coincidentally, the example in (33) also has a hesitative *imashti*, showing that *ima shuti* and *mashti* are now distinct lexical items.

- (33) *ima shuti=ta a-ra-∅ pay amu kwinta-u-n=mi ña*
what name=INT COP-PST-3SG 3 owner tell-PROG-3SG=EV.S DSC.PRT

imashti *chasna* *sukta* *wata-yuk* *a-sha* *pay=ga* *kasna=lla*
 HES like.that six year-PROP COP-SS 3=TOP like.this=LIM

kwinta-g *a-ra-*Ø
 tell-SBJ.NMLZ COP-PST-3SG

“what was [his] name?” [we asked] “he says [that] he is the master” so um, like that, being six years old he would just tell [stories] like this’
 [lwc_qvz001:25]

Thus, it is reasonable to conclude that the *imasti* in Cordero’s (1892: 40) dictionary derives from *ima shuti* ‘what name’, which – given its lexical semantic content of asking for a name – is roughly equivalent to the English ‘whatchamacallit’ and makes for a prime source from which a placeholder can be derived.

Cordero’s dictionary also has an entry for a deverbal *imashutina*,²⁹ which indicates that at least as far back as the nineteenth century *ima shuti* ‘what name’ could serve as a verbal placeholder in addition to a nominal placeholder. The given definition “doing that for which I do not remember the name of in this instant” makes the placeholder function of *imashutina* obvious but does not indicate if it was used as a pro-verb.

Concerning the historical derivation of the different functions of *mashti*, the lexical phrase *ima shuti* ‘what name’ clearly represents the starting point from which *mashti* and its functions derived. Lexicalized constructions involving an interrogative and a noun are a common source of placeholders (Podlesskaya 2010: 13). From this starting point, *ima shuti* was phonetically reduced and took on new functions: placeholder filler, independent placeholder, hesitative filler, and pro-verb. A diagram representing this process is given in Figure 13. The independent placeholder and placeholder filler functions are rolled into a single “placeholder” category for this discussion.

At the outset, the *ima shuti* ‘what name’ construction was likely employed in “whatchamacallit” type contexts of disfluency and underwent a process of phonetic erosion (as evidenced by the loss of the /u/ vowel). Semantic bleaching was likely another factor as the propositional content (the interrogative force and the “name” noun) was lost over time.

As a watchamacallit type placeholder, *mashti* then took on new functions as a hesitative filler and pro-verb. To derive hesitative fillers, Heine (2013) proposes a process of “cooptation” whereby lexical content is “co-opted” and redeployed at the “thetical” level of discourse (i.e. not syntactically integrated). Thus, *mashti* –

²⁹The -na suffix is an infinitive nominalizer in the Colombian-Ecuadorian varieties of Quechua and is often used as the ‘citation form’ for a given verb.

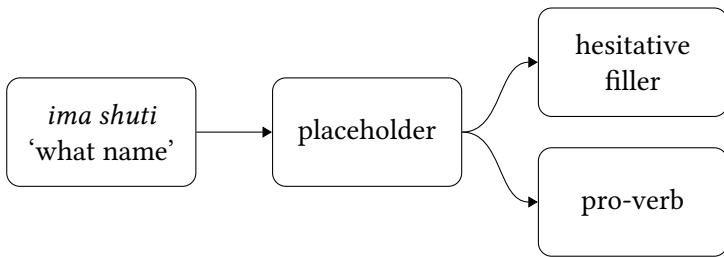


Figure 13: Proposed development of *mashti* functions in Northern Pastaza Kichwa

as a placeholder filler commonly deployed in situations of hesitation and disfluency – would have taken on a new role as a more general hesitative filler without needing to be syntactically integrated into the utterance.

For the pro-verb use of *mahsti*, I propose that the placeholder filler *mashti* is undergoing a process of (re)lexicalization. As I show in §3.3 and §4.3, the pro-verb *mashti* is less prone to occurring in contexts of disfluency and it does not target a yet unuttered referent. Instead, pro-verb *imashti* makes anaphoric reference to a previously uttered predicate event similar to the English ‘do’. This represents an instance of lexicalization, wherein a new meaning that is paired with the form *mashti* and is readily interpretable as a holistic unit (in this case, a verbal root) (Lehmann 2002). Crucially, the new pro-verb meaning is more semantically specific than that of a placeholder. Lexicalization “repackages” a form with a new meaning. The new meaning is still semantically “light” as is the case with pro-forms and manner demonstrative verbs (Guérin 2015), but it is somewhat more filled-in compared to a placeholder.

This brief discussion on the development of the different functions of *mashti* represents cursory speculation on my part. A more dedicated diachronic analysis of the development of *mashti* and fillers in general would make for a fruitful direction for future work.

6 Summary and conclusions

In this chapter I have shown how *mashti* is used in Northern Pastaza Kichwa as a hesitative filler, placeholder filler, and resumptive pro-verb. Each category can be defined by their function and how they interact with surrounding discourse. Both the hesitative and placeholder filler uses of *mashti* signal that the speaker is experiencing a disfluency event and serve to buy time to reorient their position in discourse. The hesitative filler *mashti* is not syntactically integrated into its

surrounding utterance. The placeholder filler *mashti*, by contrast, is syntactically integrated into the surrounding utterance and mirrors the target lexical item that is eventually produced when the disfluency has been resolved. The pro-verb *mashti* is not a filler. Instead, it targets a previously uttered predicate to reiterate and remind the interlocutor of a previously mentioned event.

I have also shown that there are ambiguous cases wherein the morphosyntactic properties of a given instance of *mashti* may not be enough to identify its function. To this end, this analysis considers multimodal phenomena to help classify morphosyntactically ambiguous instances of *mashti*. I show that multimodal elements such as pauses, false starts, prosodic lengthening, gaze aversion, and gesture are important correlates in identifying disfluency episodes, and I raise the possibility that such elements could be used to disambiguate the different functions of *mashti* in a larger dataset. I have also discussed how *mashti* likely arose from a lexical phrase and speculated on the developmental pathways that gave rise to the various extant functions of *mashti* in Northern Pastaza Kichwa.

There remain many outstanding questions and avenues of future investigation. First, a more careful consideration of the prosodic characteristics of disfluency such as the relevance of intonational unit boundaries and the lengths of hesitations and pauses might be useful in determining the function of uninflected instances of *mashti*. Second, a larger corpus with more tokens of placeholder and pro-verb uses would be useful, given that 75% of the tokens fall into the category of hesitatives in the present corpus. Third, there is some circular logic in my analysis, in that I analyzed the association between the correlates of disfluency and the different functions of *mashti* while simultaneously using the correlates of disfluency to categorize the different functions of *mashti* in the first place. A better approach would involve using a subset of a larger corpus whereby unambiguous tokens of *mashti* could be categorized independently by morphosyntactic criteria and then the association between the categories and co-occurring multimodal phenomena could be better evaluated.

The development of the form of *mashti* itself also warrants further investigation. I do not have an explanation for the seemingly optional presence of the initial /i/ vowel in some cases, nor for why in some other Quechuan varieties the form is *masti* with an /s/ consonant instead of /ʃ/. It would also be useful to investigate whether *mashti* in other varieties of Quechua shows the same tripartite range of functions that it possesses in Northern Pastaza Kichwa. Additionally, since the use of *mashti* as a filler appears to be limited to varieties of the Quechua II branch, it would be worth investigating whether the Quechua I languages use it, and if not, what material is deployed as fillers in those varieties.

Finally, it is worth considering a broader view of the use of other fillers in Northern Pastaza Kichwa. In §2.3, I listed the other items that are deployed as fillers in Northern Pastaza Kichwa. It would be illuminative to investigate how often these fillers are deployed in discourse and whether or not their functions and distributions mirror those of *mashti*.

This chapter thus contributes to the typological research on fillers and place-holders by describing a multifunctional filler in an Indigenous South American language, through a novel approach that takes into account the multimodal and prosodic signals associated to its different discourse functions. As stated in §1, the investigation of “peripheral” linguistic phenomena like fillers is often ignored in language documentation and description. However, the study of said peripheral linguistic phenomena has important considerations for minority language revitalization and pedagogy. To date, minority language pedagogy has largely focused on samples of decontextualized and monologic language that show regular paradigms and predictable patterns of use. Such focus on “well-behaved” instances of language ignores the importance of metalinguistic awareness and conversation, which is the epicentre of language use (S. Rice 2021). Indigenous minority languages in particular are often situated in primarily oral contexts, wherein face-to-face interaction is the place where language “lives” (S. Rice 2017). Fillers and other “peripheral” linguistic phenomena are an essential part of interactional language use (Ameka 1992). The documentation and description thereof is thus necessary in the pursuit of pedagogical development oriented towards natural and authentic language use. In other words, learning to be “fluently disfluent” should be an essential step in a minority language pedagogy program.

Acknowledgements

This research was made possible by funding from the US Department of Education’s Foreign Language and Area Studies program (FLAS), and the Endangered Language foundation. Special thanks goes to the Amazonian Kichwa communities in Pastaza Ecuador for their time, stories, and friendship. In particular, I thank Bélgica Dagua Toqueton, Norma Ruiz Cadena, and Edwing Wilfrido Nuñez for their stellar translation work and insights. Lastly, I am grateful to the editors of this volume, Brigitte Pakendorf and Françoise Rose as well as the two anonymous reviewers of this chapter, Sally Rice, Jorge Rosés Labrada, and Simeon Floyd for their guidance, comments, and suggestions. All shortcomings of this work remain my own.

Abbreviations

1	first-person	MIR	mirative
3	third-person	NMLZ	nominalizer
ALL	allative	INT	interrogative
BEN	benefactive	LIM	limitative
CAUS	causative	O	other-perspective
COP	copula	OBJ	object
CONJ	conjunctive	PROX	proximal
DEM	demonstrative	PERF	perfective
DIM	diminutive	PH	placeholder
DIST	distal	PROG	progressive
DS	different-subject	PROX	proximal
EV	evidential	PST	past
FUT	future	PROV	pro-verb
GEN	genitive	PURP	purpose
HES	hesitative	S	self-perspective
INF	infinitive	SBJ	subject
IMP	imperative	SS	same-subject
INTJ	interjection	TOP	topic
LOC	locative	TRLC	translocative
MID	middle		

References

- Adelaar, Willem F. H. & Pieter Muysken. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.
- Albarracín de Alderetes, Lelia Inés. 2009. *La quichua: Gramática, ejercicios y diccionario castellano-quichua*. Buenos Aires: Editorial Dunken.
- Ameka, Felix. 1992. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics* 18(2-3). 101–118. DOI: 10.1016/0378-2166(92)90048-G.
- Amiridze, Nino. 2010. Placeholder verbs in Modern Georgian. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 67–94. Amsterdam: John Benjamins.
- Barbera, Joseph & William Hanna. 1947. *Catfishin'*. <https://archive.org/details/Tomandjerryliberationproject/Tom+And+Jerry+028-Cat+Fishin.mp4#>.
- Bendezú-Araujo, Raúl. 2023. Verum, focus and evidentiality in Conchucos Quechua. *Zeitschrift für Sprachwissenschaft* 42(3). 611–642. DOI: 10.1515/zfs-2023-2016.

- Betz, Simon, Nataliya Bryhadyr, Olcay Türk & Petra Wagner. 2023. Cognitive load increases spoken and gestural hesitation frequency. *Languages* 8(1). DOI: 10.3390/languages8010071.
- Brône, Geert, Bert Oben, Annelies Jehoul, Jelena Vranjes & Kurt Feyaerts. 2017. Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics* 28(3). 449–483. DOI: 10.1515/cog-2016-0119.
- Butterworth, Brian. & Geoffrey Beattie. 1978. Gesture and silence as indicators of planning in speech. In Robin N. Campbell & Philip T. Smith (eds.), *Recent advances in the psychology of language: Formal and experimental approaches*, 347–360. New York: Plenum.
- Christenfeld, Nicholas, Stanley Schachter & Frances Bilous. 1991. Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research* 20(1). 1–10. DOI: 10.1007/BF01076916.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Cordero, Luis. 1892. *Diccionario quichua-español, español-quichua*. Cuenca: Universidad de Cuenca.
- Döhler, Christian. 2025. *That* placeholder in Komnzo. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 275–313. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697589.
- ELAN (Version 6.8) [computer software]*. 2024. Nijmegen: Max Planck Institute for Psycholinguistics. <https://archive.mpi.nl/tla/elan>.
- Enfield, N. J. 2003. The definition of WHAT-d'you-call-it: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Feyaerts, Kurt, Geert Brône & Bert Oben. 2017. Multimodality in interaction. In Barbara Dancygier (ed.), *The Cambridge handbook of cognitive linguistics*, chap. 9, 135–156. Cambridge: Cambridge University Press. DOI: 10.1017/9781316339732.010.
- FieldWorks Language Explorer (Version 9.0.17.455 RC) [Computer software]*. 2021.
- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 1–9. Amsterdam: John Benjamins.
- Fox Tree, Jean E. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language* 34(6). 709–738. DOI: 10.1006/jmla.1995.1032.
- Goodwin, Marjorie Harness & Charles Goodwin. 1986. Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62(1-2). 51–76. DOI: 10.1515/semi.1986.62.1-2.51.

- Graziano, Maria & Marianne Gullberg. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology* 9(879). 1–17. DOI: 10.3389/fpsyg.2018.00879.
- Grzech, Karolina. 2016a. *Discourse enclitics in Tena Kichwa: A corpus-based account of information structure and epistemic meaning*. University of London (PhD Thesis). (Doctoral dissertation).
- Grzech, Karolina. 2016b. The non-evidential meaning of the Tena Kichwa “direct evidential”. *York Papers in Linguistics* 3. 73–94.
- Grzech, Karolina, Anne Schwarz & Georgia Ennis. 2019. Divided we stand, unified we fall? The impact of standardisation on oral language varieties: A case study of Amazonian Kichwa. *Revista de Llengua i Dret* 71. 123–145.
- Guérin, Valérie. 2015. Demonstrative verbs: A typology of verbal manner deixis. *Linguistic Typology* 19(2). 141–199. DOI: 10.1515/lingty-2015-0006.
- Hayashi, Makoto. 2003. Language and the body as resources for collaborative action: A study of word searches in Japanese conversation. *Research on Language and Social Interaction* 36(2). 109–141. DOI: 10.1207/S15327973RLSI3602_2.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Hayashi, Makoto & Kyung-Eun Yoon. 2010. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders*, 33–66. Amsterdam: John Benjamins.
- Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6). 1205–1247. DOI: 10.1515/ling-2013-0048.
- Hennecke, Inga & Wiltrud Mihatsch. 2022. Between placeholder and filler: Degrees of prosodic prominence of French *machin* and *truc*. *Pragmatics & Cognition* 29(2). 297–323. DOI: 10.1075/pc.22003.hen.
- Hoetjes, Marieke, Emiel Krahmer & Marc Swerts. 2014. Does our speech change when we cannot gesture? *Speech Communication* 57. 257–267.
- Hömke, Paul, Judith Holler & Stephen C. Levinson. 2017. Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction* 50(1). 54–70. DOI: 10.1080/08351813.2017.1262143.
- Kany, Charles. 1969. *Sintaxis hispanoamericana*. Madrid: Gredos.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: The Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan

- (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam: John Benjamins.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kisa, Yağmur Deniz, Susan Goldin-Meadow & Daniel Casasanto. 2022. Do gestures really facilitate speech production. *Journal of Experimental Psychology: General* 151(6). 1252–1271.
- Krauss, Robert M. & Uri Hadar. 1999. The role of speech-related arm/hand gestures in word retrieval. In Lynn Messing & Ruth Campbell (eds.), *Gesture, speech, and sign*, 93–116. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198524519.003.0006.
- Ladewig, Silva H. 2011. Putting the cyclic gesture on a cognitive basis. *CogniTextes* 6. 1–22. DOI: 10.4000/cognitextes.406.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 1–18. Amsterdam: John Benjamins. DOI: 10.1075/tsl.49.03leh.
- Lickley, Robin J. 2015. Fluency. In Melissa A. Redford (ed.), *The handbook of speech production*, 445–474. Malden: Wiley. DOI: 10.1002/9781118584156.ch20.
- Lickley, Robin J. 2017. Disfluency in typical and stuttered speech. *Studi AISV* 3. 373–387. DOI: 10.17469/O2103AISV000019.
- Mandel, Anne, Siiri Helokunnas, Elina Pihko & Riitta Hari. 2014. Neuromagnetic brain responses to other person's eye blinks seen on video. *European Journal of Neuroscience* 40(3). 2576–2580. DOI: 10.1111/ejn.12611.
- Muysken, Pieter. 2000. Semantic transparency in Lowland Ecuadorian Quechua morphosyntax. *Linguistics* 38(5). 973–988. DOI: 10.1515/ling.2000.018.
- Muysken, Pieter. 2009. Gradual restructuring in Ecuadorian Quechua. In Rachel Selbach, Hugo C. Cardoso & Margot van den Berg (eds.), *Gradual creolization: Studies celebrating Jacques Arends*, 77–100. Amsterdam: John Benjamins. DOI: 10.1075/cll.34.09muy.
- Muysken, Pieter. 2019. *El kichwa ecuatoriano: Orígenes, riqueza, contactos*. Quito: Abya-Yala. 436.
- Muysken, Pieter. 2021. Substrate influence in Northern Quechua languages. In Enoch O. Aboh & Cécile B. Vigouroux (eds.), *Variation rolls the dice: A worldwide collage in honour of Salikoko S. Mufwene*, chap. 6, 133–160. Amsterdam: John Benjamins. DOI: 10.1075/coll.59.06muy.
- Nakano, Tamami & Shigeru Kitazawa. 2010. Eyeblink entrainment at breakpoints of speech. *Experimental Brain Research* 205(4). 577–581. DOI: 10.1007/s00221-010-2387-z.

- Navarretta, Costanza. 2016. The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations. In *Proceedings from the 3rd European symposium on multimodal communication, Dublin, September 17–18*, 55–61. Dublin: LiU Electronic Press.
- Nuckolls, Janis B. 2014. From quotative other to quotative self: Evidential usage in Pastaza Quichua. *Pragmatics and Society* 3(2). 226–242. DOI: 10.1075/ps.3.2.05nuc. <http://openurl.ingenta.com/content/xref?genre=article&issn=1878-9714&volume=3&issue=2&spage=226>.
- Nuckolls, Janis B. 2017. *Movie 20*. <https://youtu.be/wOzfcM3ahTQ> ? si = c1fdRo0E0jB0reyV (2 February, 2022).
- Nuckolls, Janis B. 2018. The interactional and cultural pragmatics of evidentiality in Pastaza Quichua. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, chap. 10, 201–221. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198759515.013.10.
- Nuckolls, Janis B. 2020. “how do you even know what ideophones mean?”: Gestures’ contributions to ideophone semantics in Quichua. *Gesture* 19(2-3). 161–195. DOI: 10.1075/gest.20005.nuc.
- Nuckolls, Janis B. 2024. *Quechua research group [YouTube channel]*. https://www.youtube.com/channel/UCjM-OUUVz_h6ohmktY-gW0w (10 March, 2015).
- Nuckolls, Janis B. & Alexander Rice. 2015a. *GOPR0183*. <https://youtu.be/bmSebAlA6vE> (5 May, 2021).
- Nuckolls, Janis B. & Alexander Rice. 2015b. *GOPR0245*. <https://youtu.be/FIOWZP975MY> (7 July, 2020).
- Orr, Carolyn. 1991. *Vamos a aprender Quichua de Pastaza*. 1st edn., vol. 17. Quito: Instituto Lingüístico de Verano. 31. <https://www.sil.org/resources/archives/17566>.
- Orr, Carolyn & Betsy Wrisley. 1981. *Vocabulario quechua del oriente del Ecuador*. 2nd edn., vol. 11. Quito: Instituto Lingüístico de Verano en cooperación con el Ministerio de Educación Pública.
- Osorio, Sergio, Benjamin Straube, Lars Meyer & Yifei He. 2024. The role of co-speech gestures in retrieval and prediction during naturalistic multimodal narrative processing. *Language, Cognition and Neuroscience* 39(3). 367–382. DOI: 10.1080/23273798.2023.2295499.
- Ozkan, Elif Ecem, Patrick G. T. Healey, Tom Gurion, Julian Hough & Lorenzo Jamone. 2023. Speakers raise their hands and head during self-repairs in dyadic conversations. *IEEE Transactions on Cognitive and Developmental Systems* 15(4). 1993–2003. DOI: 10.1109/TCDS.2023.3254808.

- Palacios Martínez, Ignacio M. & Paloma Núñez Pertejo. 2015. "Go up to miss thingy". "He's probably like a whatsit or something". Placeholders in focus. The differences in use between teenagers and adults in spoken English. *Pragmatics* 25. 425–451. DOI: 10.1075/prag.25.3.05pal.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Ponsonnet, Maïa. 2025. Individual preferences when using placeholders: The case of Dalabon (Australian, Gunwinyguan). In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 315–340. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697591.
- Pyers, Jennie E., Rachel Magid, Tamar H. Gollan & Karen Emmorey. 2021. Gesture helps, only if you need it: Inhibiting gesture reduces tip-of-the-tongue resolution for those with weak short-term memory. *Cognitive Science* 45(1). 1–26. DOI: 10.1111/cogs.12914.
- R Core Team. 2024. *R: A language and environment for statistical computing*. Vienna. <http://www.r-project.org/>.
- Rauscher, Frances H., Robert M. Krauss & Yihsiu Chen. 1996. Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological science* 7(4). 226–231.
- Reeve, Mary-Elizabeth. 2022. *Amazonian kichwa of the curaray river: Kinship and history in the western amazon*. Lincoln: University of Nebraska Press.
- Rice, Alexander. 2019. *Ecuadorian lowland Quechua collection of Alexander Rice*. <https://www.ailla.utexas.org/islandora/object/ailla%3A269474> (9 September, 2019).
- Rice, Alexander. 2020. Northern Pastaza Kichwa (Ecuador and Peru): Language snapshot. *Language Documentation and Description* 19. 181–196. <http://www.elpublishing.org/PID/219>.
- Rice, Alexander. 2023. *Northern Pastaza Kichwa (i)mashti data*. DOI: 10.17605 / OSF.IO/G84MV. (12 December, 2023).
- Rice, Sally. 2017. The study of indigenous languages. In Barbara Dancygier (ed.), *The Cambridge handbook of cognitive linguistics*, chap. 3, 38–58. Cambridge: Cambridge University Press. DOI: 10.1017/9781316339732.004.
- Rice, Sally. 2021. Insights from cognitive semantics for language documentation. In Thomas Li (ed.), *The handbook of cognitive semantics*, 289–327. Leiden: Brill.
- Rose, Françoise. 2025. One more thing 'thing' can do in Tupí-Guaraní languages: The Teko filler. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesita-*

- tives and placeholders, 441–484. Berlin: Language Science Press. DOI: 10.5281/zenodo.1569759.
- Rossano, Federico. 2012. Gaze in conversation. In Jack Sidnell & Tanya Stivers (eds.), *The handbook of conversation analysis*, chap. 15, 308–329. Chichester: Wiley-Blackwell. DOI: 10.1002/9781118325001.ch15.
- SayMore* (Version 3.2.21) [computer software]. 2022. <https://software.sil.org/saymore/>.
- Schachter, Paul & Timothy Shopen. 2007. Parts-of-speech systems. In Timothy Shopen (ed.), *Language typology and syntactic description: Clause structure*, 2nd edn., vol. I, chap. 1, 1–60. Cambridge: Cambridge University Press.
- Seifart, Frank. 2008. On the representativeness of language documentation. *Language Documentation and Description* 5. 60–76.
- Seraku, Tohru. 2022. Interactional and rhetorical functions of placeholders: A relevance-theoretic approach. *Journal of Pragmatics* 187. 118–129. DOI: 10.1016/j.pragma.2021.10.023.
- Seyfeddinipur, Mandana. 2006. *Disfluency: Interrupting speech and gesture*, vol. 39. Nijmegen: Radboud Universiteit Nijmegen. <https://hdl.handle.net/2066/41397>.
- Stivers, Tanya & Jack Sidnell. 2005. Introduction: Multimodal interaction. *Semiotics* 156. 1–20. DOI: 10.1515/semi.2005.2005.156.1.
- Swanson, Tod Dillon. 2017. *Luisa cadena*, "first woman sends her children away to become animals.". <https://youtu.be/w8NfgNvqCTw?si=wPkN0GMtSjPp48lb> (5 May, 2021).
- Swanson, Tod Dillon. 2020. *Bélgica Dagua*: "The strangler fig loved a girl." <https://youtu.be/Q6bYzbGCOJQ?si=LI3qZInzGSGuptyr> (5 May, 2023).
- Swanson, Tod Dillon. 2021. *Delicia*, "noah and the flood.". <https://youtu.be/BNGxuKfVdQ0?si=iUp1cB1rA3rVTxOE> (3 March, 2022).
- Swanson, Tod Dillon. 2024. *Iyarinapamba* [YouTube channel]. <https://www.youtube.com/user/iyarinapamba> (9 September, 2019).
- Sweetser, Eve & Kashmiri Stec. 2016. Maintaining multiple viewpoints with gaze. In Barbara Dancygier, Wei-lun Lu & Arie Verhagen (eds.), *Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities*, 237–258. Berlin & Boston: De Gruyter Mouton. DOI: 10.1515/9783110365467-011.
- Thompson, Sandra A. & Ryoko Suzuki. 2014. Reenactments in conversation: Gaze and recipiency. *Discourse Studies* 16(6). 816–846. DOI: 10.1177/1461445614546259.
- Torero, Alfredo. 1964. Los dialectos quechuas. *Anales Científicos de la Universidad Agraria* 2(4). 446–478.

- Vallejos Yopán, Rosa. 2023. From demonstrative to filler: *Este* in Amazonian Spanish and beyond. *Linguistics* 61(3). 651–678.
- Warren Carney, Lisa. 2016a. *Chingarishka turi*. <https://youtu.be/UjO076a0FP8?si=oFuaHFaihokry8FW> (6 March, 2021).
- Warren Carney, Lisa. 2016b. *Manga allpa muskuy*. <https://youtu.be/CX2KjggHQ-k?si=IbDlsDiCaefichb5T> (9 May, 2022).
- Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. <https://ggplot2.tidyverse.org>.
- Zorraquino, Martín, María Antonia Portolés & José Portolés. 1999. Los marcadores del discurso. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 4051–4213. Madrid: Espasa Calpe.

Appendix A Summary of recordings in corpus

Consultants: BD: Bélgica Dagua, DD: David Dagua, DE: Delicia Dagua, ED: Eu-lodia Dagua, LC: Luisa Cadena, SV: Sisa Viteri, TD: Teolinda Dagua

Register: C: Conversation, N: Narrative, SD: Stimulus description,

Collector(s): AR: Alexander Rice, Janis Nuckolls, LWC: Lisa Warren Carney, TS: Tod Swanson

ID	Consultant	Duration (mm:ss)	Register	Collector(s)	Source
jbn_qvz002	LC	15:27	N	JN	(Nuckolls 2017)
jbn_qvz003	LC	14:08	N	JN	(Nuckolls & Rice 2015b)
jbn_qvz005	LC	06:29	N	JN	(Nuckolls & Rice 2015a)
lwc_qvz001	BD	05:32	N, C	LWC	(Warren Carney 2016a)
lwc_qvz002	ED	02:58	N, C	LWC	(Warren Carney 2016b)
qvz001	ED	13:06	N, C	AR	Rice (2019)
qvz002	BD	11:18	N, SD	AR	Rice (2019)
qvz007	BD	14:40	N	AR	Rice (2019)
qvz008	BD	06:38	N	AR	Rice (2019)
qvz010	BD	15:05	N, SD	AR	Rice (2019)
qvz011	ED	05:54	N	AR	Rice (2019)
qvz014	ED	05:43	C	AR	Rice (2019)
qvz015	BD	09:51	C	AR	Rice (2019)
qvz017	BD	02:18	C	AR	Rice (2019)
qvz018	BD	01:06	C	AR	Rice (2019)
qvz020	TD	07:55	N	AR	Rice (2019)
qvz022	TD	15:43	N, SD	AR	Rice (2019)
qvz026	BD	14:04	N, C	AR	Rice (2019)
qvz042	SV	23:06	C, N	AR	restricted access
tds_qvz003	BD	06:27	N	TS	(Swanson 2020)
tds_qvz010	ED	02:36	N	TS	(Swanson 2021)
tds_qvz023	LC	09:53	N	TS	private data
tds_qvz024	LC	03:12	N	TS	private data
tds_qvz025	LC	11:41	N	TS	(Swanson 2017)
tds_qvz031	DD	15:28	N	TS, AR	private data
tds_qvz032	DD	26:22	N	TS, AR	private data
tds_qvz033	DD	07:47	N	TS, AR	private data
tds_qvz034	BD	05:30	N	TS, AR	private data
tds_qvz035	BD	05:47	N, C	TS, AR	private data
tds_qvz036	BD	05:19	N	TS, AR	private data
tds_qvz037	BD	33:14	N, C	TS, AR	private data
tds_qvz039	ED	04:31	N	TS, AR	private data
tds_qvz044	DE	15:31	N	TS	(Swanson 2021)
tds_qvz045	BD	13:52	N	TS	private data

Chapter 12

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

Françoise Rose

Laboratoire Dynamique du Langage (CNRS & Université Lumière Lyon 2)

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 van der Auwera & Krasnoukhova (2021: 90) 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 **ma?*e 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 **ma?*e ‘thing’ has been reconstructed for Proto-Tupí-Guaraní (Rodrigues 1984: 36; Jensen 1998: 536; Souza Mello 2000: 230). Two papers focus on the uses of the noun ‘thing’ and its grammaticalized versions in the Tupí-Guaraní language group: Dietrich (1994) and van der Auwera & Krasnoukhova (2021). A list of these uses follows, and most of them will later be discussed and illustrated with Teko data (§2).

- General term for any non-human entity
- Incorporated in a verb with an intransitivizing effect
- Used in *ba?*e-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 **ma?*e 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 (Jensen 1999; Rose 2023). Nevertheless, Dmitry Gerasimov has pointed out that the noun for ‘thing’ with some additional morphology *ma?*e-*rā* has a “whatchamacallit use” in Paraguayan Guarani, 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 **ma?*e ‘thing’ (van der Auwera & Krasnoukhova 2021).

The filler use of the reflex of **ma?*e has also already been reported in the first reference grammar of Teko (Rose 2011), and is the topic of the present study, which is inspired by the literature on the richness of reflexes of **ma?*e 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. (van der Auwera & Krasnoukhova 2021: 90)

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 (Navet 1994). 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 (Rodrigues 1984; Rose & Drude Submitted). 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 (Maurel 1998), and a first reference grammar was published a decade ago (Rose 2011). 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 (Rose 2011: 176). 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-irur o-mumup=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) kito-r-ehe e-iba.*
 EXIST child-TRANSF HES frog-RELN-with 3-pet
 ‘There is a child with his um... pet frog.’ [13.001]

- (3) *Kor kito-r-aʔir* (0.8) *o-ipuŋ* (0.7) ***baʔe-pope-dʒi*** (2.4)
then frog-RELN-son 3-put thing-in-LOC
bokar-a-pe-dʒi *o-ipuŋ*.
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 (Podlesskaya 2010; Hayashi & Yoon 2006). *Thingy* is a prototypical example of a placeholder in English (Palacios Martínez & Núñez Pertejo 2015), and *uh* of a hesitator (Clark & Fox Tree 2002).

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 (Chousou-Polydouri et al. n.d.). 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. Grenand 1989: 267, on Wayampi), grammatical descriptions (e.g. Godoy 2020, on Ka'apor) and corpora (e.g. Gasparini 2015, 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 van der Auwera & Krasnoukhova (2021: 90) 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 (Rose 2018) 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-Lescure et al. 1987; Ti'iwan et al. 1993; Maurel 1991; Maurel 2000) and one leaflet introducing the euro currency (Association Solidarité Guyane 2000). 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 ELAN (2022), 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

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

²When 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.

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.

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 §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 (Boersma & Weenink 2023) 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 mor-

³Note that the initial plosive may be prenasalized [mb], and that the glottal stop phoneme may be produced without a closure of the vocal folds (Rose 2021: 4–6). In this case, this phoneme is not segmented as an interval.

phology) 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 §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 <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.

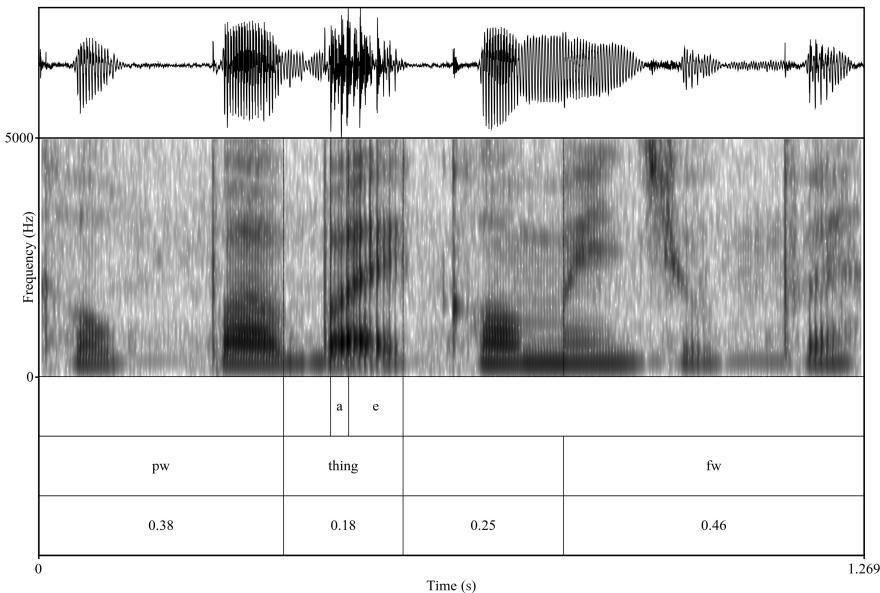


Figure 1: Praat analysis of an occurrence of *baʔe* as a noun

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.

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

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 occurrences 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, §2 will present the uses of *baʔe* in Teko. Then, §3 will analyze some uses of *baʔe* as a hesitator, which in synchrony cannot be analyzed as a noun anymore. §4 will later explore the discourse uses of the *baʔe* noun ‘thing’. Finally, §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 (Dietrich 1994; Jensen 1998; van der Auwera & Krasnoukhova 2021), the same form is employed in the language for nominalization (§2.2) and

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

interrogation (§2.3), but not for negation. Beside these well-described uses of the reflexes of **maʔe*, Teko *baʔe* is also used as a verb (§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 §2.1, in nominalization in §2.2, in interrogation in §2.3, and as a verb in §2.4. This background information is necessary to later determine the part of speech of the *baʔe* elements under study in §3 and §4, and to discuss diachronic paths of evolution in §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-apig=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-pihig!**
 thing EXH-1SG-catch
 ‘I want to catch them (girls who later happen to be mermaids).’ [32.043]

- (6) *Apam-a-baʔe-pe=āhā node-r-apidz-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-wirakoti.*
 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-apig=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-pihig!*
 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 (Cachine et al. 2020: 101), 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 §4.3).

- (10) *Si-ro-nan aŋ baʔe panaisi wāiwi-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 (Cutting 2019). 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 context, or on the exophoric context to identify their referent (Halliday & Hasan 2013; Adler & Moline 2018). General nouns participate in vague expression, i.e.

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

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” (Cutting 2007: 4). And indeed, tokens of the noun *baʔe* in spontaneous 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 *ba?e* are specific and identifiable, either anaphorically, cataphorically or exophorically. In (7), from a picture description task, the speaker makes reference to a boucan⁷ 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. *Ba?ekom* refers back to the globality of these individual referents. Example (9) is from a text with two female heroes to which *ba?e* refers.

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

- (11) “Ware?ete-a?u a-ma?ē ba?e-kom-a-r-ehe pe-r-upi
 be_beautiful-TOP_SW 1SG-see thing-PL-REF-RELN-to path-RELN-along
 a-ho-nam”, e?i-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 *ba?e* 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 *ba?e*, 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 Rose (2011: 176). 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 *ba?e*, or also *ba?eza?u* ‘food’⁹ after verbs of cooking and ingestion.

⁷A ‘boucan’ is a wooden frame on which meat or fish is smoked.

⁸See also discussion and (1) in §1.2: it is precisely in the section of the grammar (Rose 2011) which explains this syntactic requirement that the issue of fillers was first mentioned.

⁹The nominal stem *baʔe-zaʔ-u* ‘food’ is made of *baʔe* ‘thing’, *za*, which is formally similar to the indeterminate subject prefix, and the verb *ʔu* ‘ingest’.

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

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-potir*
kunami-flower
'kunami flower (type of creeper)'
b. *baʔe-potir*
thing-flower
'flower (general term)'
- (14) a. *basakara-r-upiʔa*
hen-RELN-egg
'chicken egg'
b. *baʔe-r-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" (Hennecke & Mihatsch 2022: 300). I therefore do not analyse *baʔe* in (12) to (14) as a placeholder. Its actual placeholder uses will be discussed in §4.2.

¹⁰See Rose (2011: 28; 161) on Teko obligatorily possessed nouns and their person prefix *zo-* for non-specific human possessors.

¹¹Nevertheless, "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." (Fronek 1982: 643)

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-wər*
 thing-pass-FREQ
 'beads string reminding the order of traditional chants'

[Cachine et al. 2020:15]

Incidentally, the productive Teko relativizer/complementizer *=māʔē* (Rose 2011: 343–354) has been reconstructed as a clause nominalizer **βaʔe* in Proto-Tupí-Guaraní (Jensen 1998: 595), and this reconstruction is hypothesized to be linked to **maʔe* ‘thing’ by van der Auwera & Krasnoukhova (2021: 82). 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 *baʔekʷər* in (16).

- (16) *Aŋ baʔekʷər*¹³ [a-mebeʔu-tar-a=**māʔē**].
 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)kar=iŋ, e-pari-kom?*
 thing=INTER 2PL-search=CONT 1SG-grandchild-PL
 'What are you looking for, my grandchildren?'

[12.114na]

¹²The other cases are *baʔe-wa* ‘restaurant’ where *wa* is a non-productive verbal root for ‘eat’, *baʔe-r-ai* ‘illness’ where *ai* ‘be in pain’ is not a verb but a type of adjectival predicate (called “attributive” in Rose 2011: 39–51), and *baʔe-za-?u* ‘food’ (see footnote 9).

¹³This term results from the lexicalization of *baʔe* ‘non-human’ and the suffix *-kʷər*, which is often translated in English with the prefix ‘ex-’.

- (18) *Baʔe-ne=eʔe* *ta=aŋ* *o-baʔe? awa ta=aŋ* *o-ig?*
 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. van der Auwera & Krasnoukhova (2021: 74) actually relate the latter root to **maʔe*.

Table 2: Interrogative forms and their likely sources

interrogative form	gloss	source	source gloss
<i>awa</i>	‘who?’		
<i>baʔe</i>	‘what?’		‘thing’
<i>manan ~ manani</i>	‘how?’	<i>ma+nan</i>	Q+‘so’
<i>manam</i>	‘when?’	<i>ma+nam</i>	Q+‘when’
<i>mati</i>	‘where to / from?’	<i>ma+ti</i>	Q+towards
<i>mananāhā</i>	‘how many?’	<i>ma+nan+āhā</i>	Q+‘so’+‘only’
<i>may</i>	‘which?’	<i>ma+aŋ</i>	Q+DEM
<i>baʔamō</i>	‘what for?’	<i>baʔe+am(ō)</i>	‘thing’+TRANSF
<i>baʔe-r=ehe</i>	‘why?’	<i>baʔe-r=ehe</i>	‘thing’+RELN+POSTP
<i>baʔe-wi</i>	‘from what?’	<i>baʔe-wi</i>	‘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 §2.1.

Semantically, *baʔe* in questions is used for non-human participants (translated as ‘what’ in the dictionary (Cachine et al. 2020: 14)). 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) *Baʔe=sipo zewe e-me-menō?*
 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 (Rose 2011: 27). 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) *Bedžu (0.3) o-baʔe (0.3) zapēhē-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 (Guérin 2015) when introducing ideophones, which depict events, i.e. actions or states, in Teko (Rose 2024). This light verb use is illustrated in (25).

- (22) *Aipo=po a-baʔe-tar?*
 now=INTER 1SG-do-FUT
 ‘Will I do something now?’ [01.007]
- (23) *Ani (1.4), ani, mamen nan ere-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-ra.*
 1SG-do-FUT-PL.O 1SG-say-PFV
 ‘I told you that I will do (it to) them.’ [21.205]
- (25) *Moj o-baʔe.*
 IDEO.be_dark 3-make
 ‘S/he turns the light off.’ [elicited]

When introducing an ideophone, *baʔe* contrasts with the partially synonymous 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) *Moj 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* (Rodrigues 1984: 36; Souza Mello 2000: 235–236). Souza Mello (2000: 236) 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” (Hayashi & Yoon 2006: 507). Examples of hesitators are *um* and *uh* in English (Clark & Fox Clark & Fox Tree 2002), and *este* in Amazonian Spanish (Vallejos Yopán 2023). 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 *ba?e* 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 *ba?e* 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 *ba?e* to be presented in §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 *ba?e*, i.e. the “delayed constituents” in the terminology of disfluency analysis.

Table 3: Type of delayed constituent following *ba?e* 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 *ba?e* are most often noun phrases (some within a postpositional phrase), but also sometimes verbs or other categories. In the latter case, *ba?e* appears in clause-initial position. Examples follow in (27) to (29).

- (27) *Nan-am-a=ko=p* *mamā baʔe* (2) *nana=ne*
 thus-TRANSF-REF=INJ=CONT mommy HES pineapple=CONTR
o-boʔi-i.
 3-split-DIM
 ‘At that time, Mum was cutting into pieces the um... pineapple.’ [21.134]
- (28) *Min-a=we=zepe,* *baʔe* (0.7) *oʔu-o-ʔu wasei.*
 formerly-REF=also=CONCES HES RED-3-eat açai
 ‘In the old days, um... they were often eating açai.’ [31.042]
- (29) *Baʔe mokoj wā̄wī-kom* (0.9) *o-iba=we o-menō.*
 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) *pitan-am* (2.3) *baʔe* (1.2) *kito-r=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-koti* (0.8) *tebesi-koti-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) *Mataʔere¹⁴* *baʔe de-zar-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.

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

- (33) **Ba?e** (0.2) *zawar-a-kom-a-t=ehe=ãhã=pi?ia=terā o-ma?ẽ.*
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 *ba?e* as nouns is their prosody. When used as a hesitator, *ba?e* is generally lengthened and preceded and followed by a pause. This is illustrated in Figure 2, representing (34).

- (34) *Kiwo=ne pe-itfi-itfig* (0.2) **ba?e** (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]

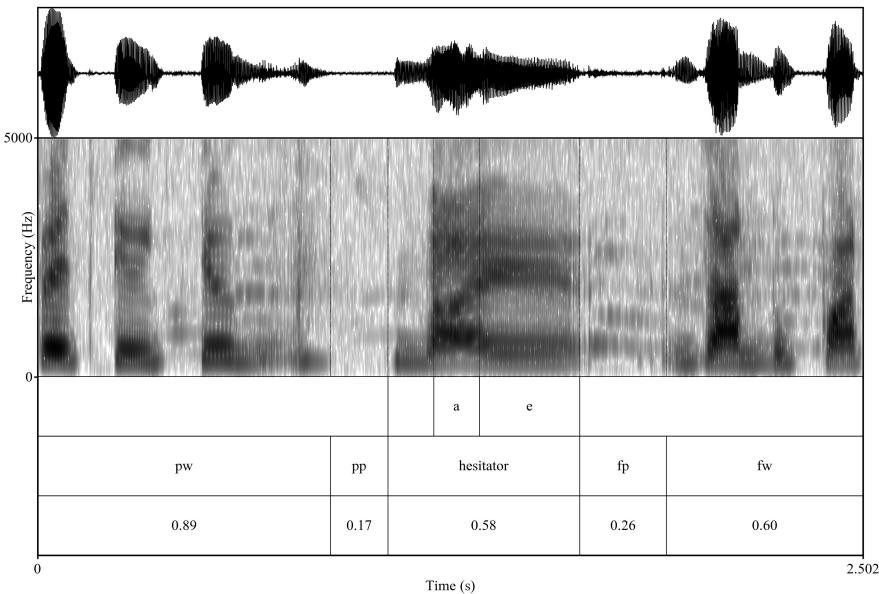


Figure 2: Praat analysis of an occurrence of *ba?e* as a hesitator

The length of *ba?e* 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 *ba?e* (in the 11 occurrences without any morphology) and its hesitator use.

Table 4: *ba?e* final lengthening

Functions of <i>ba?e</i> (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 *ba?e* is much more frequent when it is a hesitator rather than a noun. Additionally, in the cases where *ba?e* 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 *ba?e*, when there is a pause (third column of Table 5).

Table 5: Pauses before and after *ba?e*

Functions of <i>ba?e</i> (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 *ba?e* 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 (Hayashi & Yoon 2006: 525; Vallejos Yop n 2023: 24). Within the Teko linguistic system, “hesitator” could be considered a separate part of speech.

4 Discourse uses of the noun *ba?e*

This section proposes that the noun *ba?e* 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 §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” (Mauri & Sansò 2018: 2).

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 (Overstreet & Yule 2021: 1–2; 40–41).

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. (Mauri & Sansò 2018: 12)

Two examples are given as illustrations:

- (35) *Ikemin-a-nam* (0.5) *za-sinjar*, *za-wera*, *wiwa-pe pira*
 old_days-REF=SUB INDET-poison_fish INDET-hunt arrow-with fish
za-džika (0.8), *di-za-mobor-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 arakapusa* (0.8), *dati fort, t-iru,* ***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 implicitly 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 construction is specified by a relative clause, a structural option known for inserting more specific information about the ad hoc category (Overstreet & Yule 2021: 40).

- (37) *O-pihig tikadir o-?oram idža baʔe ai-māʔē-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 translated 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 probably 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 *idža* ‘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, *ba?e* 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 *ba?e* 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 *ba?ekom* immediately follows *tiru*, and is not lengthened. The long pause is an interclausal pause.

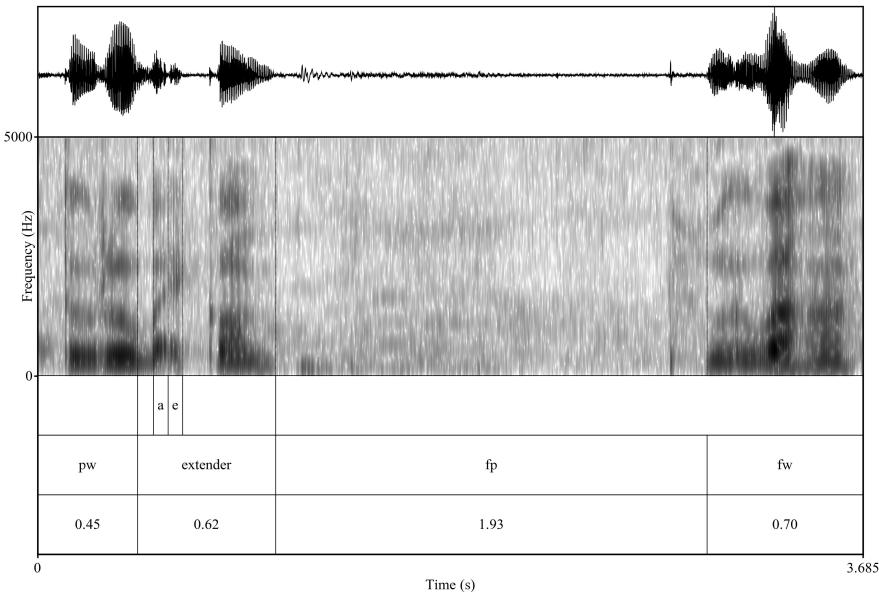


Figure 3: Praat analysis of an occurrence of *ba?e* as an extender

Even though Mauri & Sansò (2017: 65–66) 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 *ba?e*. It has been explained that this noun might have either a specific or a non-specific referent in its basic use (see §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 (Rose 2011: 146–150),¹⁵ nor nouns for ‘type, gender, sort, class, category’ (Cachine et al. 2020). 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 Rose (2012: 55–56), 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 mamā-kom-a-puri a-džu=j.*
Dad Mum-PL-REF-at 1SG-be=CONT
'I am at Mum and Dad's.' [22.027]
- (39) *I-ji-kom* (0.2) "m" e?i.
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).

¹⁵The conjunction *?oram* ‘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).

¹⁶The two clear examples of the associative plural use are on kin terms, which is cross-linguistically expected (Daniel & Moravcsik 2013). The associates to the focal member are other family members.

¹⁷There 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). (Overstreet & Yule 2021: 1)

- (40) *Kor kito-r-aʔir* (0.8) *o-ipuŋ* (0.7) ***baʔe-pope-dʒi*** (2.4)
 then frog-RELN-son 3-put thing-in-LOC
bokar-a-pe-dʒi *o-ipuŋ*.
 jar(Fr.)-REF-in-LOC 3-put
- 'Then he puts the baby frog in a thingy, in a jar.' [13.003]

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, *bokar*, 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 Fox (2010: 2), 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) *Kir* (1.3) *tou baʔe-pori-ne* *o-ho o-ʔa*, (0.5) *kaʔiboʔi-pori*.
 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, Podlesskaya (2010: 18) 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 (Rose 2011: 398–399).¹⁸ As a second-position clitic (Rose 2011: 395–396), 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

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

- 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).

Seraku (2022: 123) 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” (Schegloff 1979: 274), 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 (Seraku 2022: 122).

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,

¹⁹In 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.

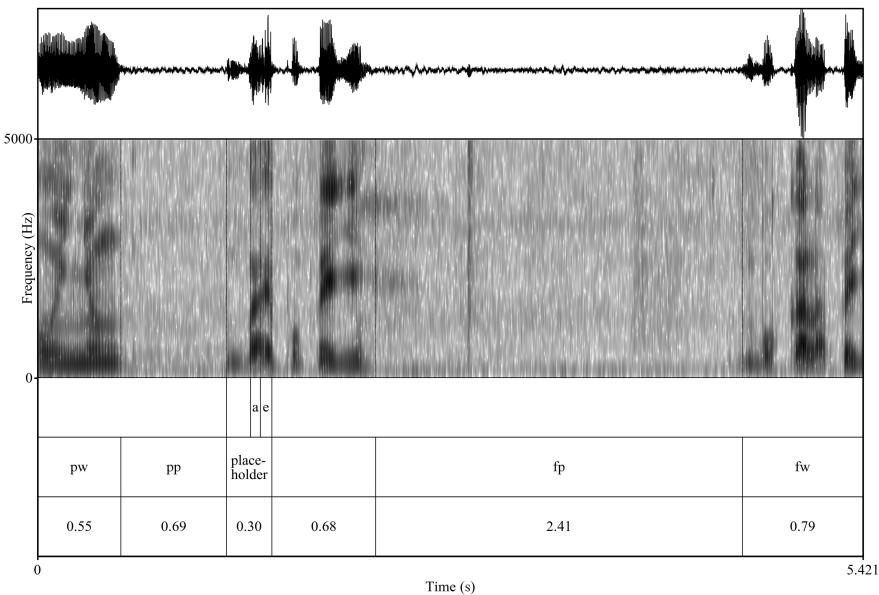


Figure 4: Praat analysis of an occurrence of *ba?e* as a placeholder

the placeholder is used to create suspense (Luelsdorff 1995) by postponing a bit the naming of the living entity under focus.

The question remains open whether *ba?e* 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 me a story. Her request was not followed, so it is unclear what character was targeted. Still, *ba?e* here is possibly a placeholder, uttered to avoid the name of a dispreferred (maybe threatening) entity.

- (42) *A?e ba?e d-o-aihi-dži-ma?ē-pe ere.*
DEM thing NEG-3-like-NEG=REL-about 2SG.say
‘Tell (the story) about the **one** we don’t like.’ [31.024]

When *ba?e* is not followed by a target, it might not always be possible to make a straightforward distinction between the uses of *ba?e* 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).

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-ʔar-a=itʃe kor (0.2) baʔe dʒaiwər-a-r-aʔiɾ.*
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 §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 (Rose 2011: 156–158). 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).

²⁰For a similar difficulty with *thing* in English, see Palacios Martínez & Núñez Pertejo (2015).

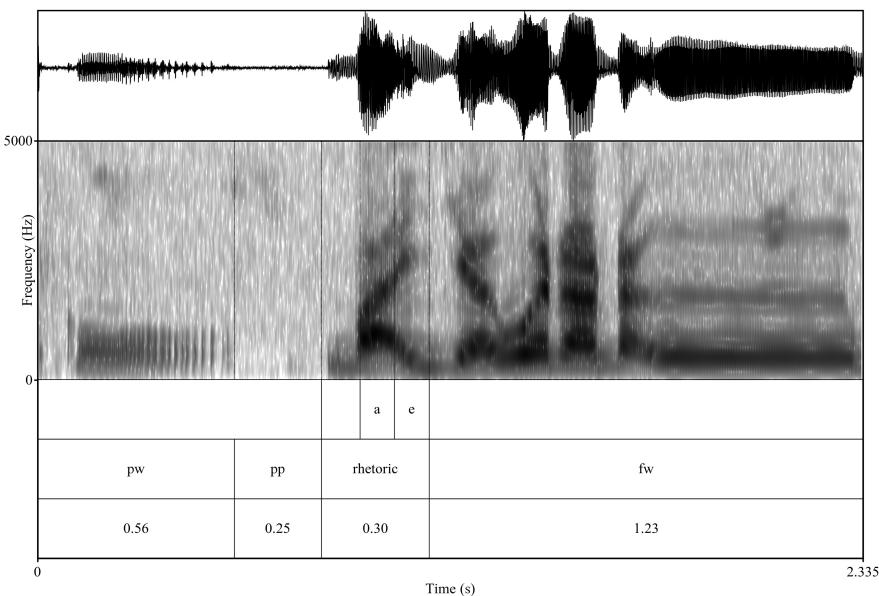


Figure 5: Praat analysis of an occurrence of the rhetorical use of *ba?e+N*

- (44) a. *pitan awak^{wər}*
 child man
 'the boy'
 b. *teko abir*
 Teko deceased_person
 'the late Tekos'

The difference is that in regular $N_{\text{head}}+N_{\text{modifier}}$ sequences, both nouns contribute semantically to the identification of the referents, while in all but one *ba?e+N* sequences, see 45, *ba?e* is not informative because the referents of the specific nouns are unquestionably 'non-human'. This raises the question whether this use of *ba?e* 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: *ba?e* 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

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” (Enfield 2003: 106), 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 *koc* ‘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 *dʒaiwər-a-r-aʔiɾ* is much lengthened, putting a very clear emphasis on this term.

As mentioned at the beginning of this paper (§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-ro-nan aŋ baʔe panaisi wāiwi-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 *panaisi*: 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

²¹A definition of “euphemism” is the following: “A euphemism is used as an alternative to a dispreferred expression” (Allan & Burridge 1991: 11).

²²Brigitte 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” (Zikmundová 2013: 106–107).

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

been recognized that fillers might have a euphemistic function, possibly linked to taboo avoidance (Enfield 2003; Cheung 2015; Palacios Martínez & Núñez Pertejo 2015). 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 (Kleiber 1987). It is well known that “general-for-specific” is a common metonymic strategy for taboo items (Burridge 2006).

Lexical substitution or metaphorical periphrasis is often employed in languages spoken in the Amazon basin, due to taboo avoidance (Dienst & Fleck 2009; Mihas 2019; Wojtylak 2015; Wojtylak 2019). This is in line with the Amazonian cosmology (de Castro 1998; Santos-Granero 2007), 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 *ba?e* 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 (Seraku 2024: 13).

What remains to be determined is whether this use results in a “constructional change” (Traugott & Trousdale 2013), i.e. on the basis of an existing noun phrase structure with two nouns, potentially a semantic bleaching of *ba?e* (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 *ba?e* elements and their use in Teko from a diachronic perspective. It starts in §5.1 by putting forward the nominal source of the hesitator within a rich network of grammaticalizations issued from the noun

²⁴In fact, the concept of ‘spirit’ is found as a secondary or a new meaning for reflexes of Proto-Tupí-Guaraní **ma?e* in some Tupí-Guaraní languages (Chousou-Polydouri et al. n.d.)

‘thing’. In §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 §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 Podlesskaya 2010; Enfield 2003). They are briefly presented below:

- pronouns: third person, demonstrative, interrogative, indefinite or emphatic, such as Japanese *ano* (Hayashi & Yoon 2006: 507).
- clauses, often with an interrogative word and naming nouns or verbs, such as English *whatchamacallit* (Enfield 2003).
- a general noun; examples from Podlesskaya (2010: 13) 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’ (Dimock 2010: 209) and Lao *qan⁰-nan⁴* ‘that thing’ (Enfield 2003).
- 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’ (Ganenkov et al. 2010: 102).

General (or generic) nouns are known as good sources for grammaticalization. In their *World Lexicon of Grammaticalization*, Kuteva et al. (2019: 432–435) list five grammaticalization targets of the noun ‘thing’: complementizer, indefinite pronoun, attributive possession marker, nominalizer, and interrogative pronoun. And indeed, van der Auwera & Krasnoukhova (2021) have pointed to a number of grammaticalization paths from the noun **ma?e* ‘thing’ in Tupí-Guaraní languages. The general picture of that network of changes is schematized in Figure 6.²⁵

²⁵The 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)” (van der Auwera & Krasnoukhova 2021: 89). The question mark after “nominalizer” is present in the original figure and represents remaining issues in the diachronic analysis (van der Auwera & Krasnoukhova 2021: 82–87).

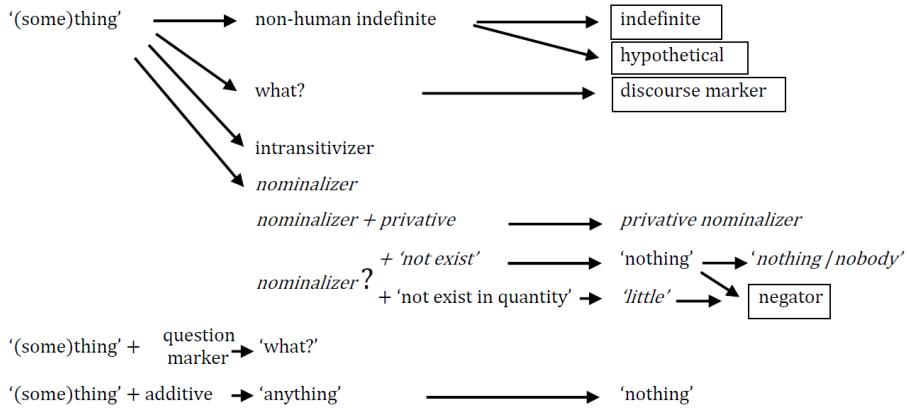


Figure 6: Diachronic developments from the noun ‘thing’ in Tupí-Guaraní languages (copied from van der Auwera & Krasnoukhova 2021: 90)

In this figure, discourse markers refer to Nheengatu clause-final ‘protest’ *ba?*, Siriono *ba* surprise particle, Warázu *ma?e* ‘yes’, Mbyá Guarani *mba?e* particle used to either introduce a proposal with a ‘how about’ meaning, or render ‘for example, perhaps’ (van der Auwera & Krasnoukhova 2021: 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 *ba?e* forms described in §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.

First, there is no morphological, syntactic, or semantic reason to suspect the nominalizing use of *ba?e* 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 §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 van der Auwera & Krasnoukhova (2021) 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 *ba?e* 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 *ba?e*. 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 *ba?e* 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 Podlesskaya (2010: 26) 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 *ba?e* but systematize its potential as a vague expression, which is not systematically implemented in its basic nominal use (see §2.1). As Hennecke & Mihatsch (2022: 300) 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 *ba?e*, 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 pragmatalization. Pragmatalization 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 metacom-municative, discourse interactional meaning” (Mihatsch 2006a: 397). 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 Diewald 2011: 373).²⁶

Hennecke & Mihatsch (2022: 304) 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 (Rice 2025 [this volume]) and Amazonian Spanish (Vallejos Yopán 2023).

In the process of pragmatalization, the types of target of *ba?e* widened from non-human nouns to constituents of all sorts and length (see Mihatsch 2006b: 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.

²⁶This 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 (Podlesskaya 2010), as not being discourse markers (Vallejos Yopán 2023), or as grammatical items (Kirjavainen et al. 2022), respectively. Second, the literature on pragmatalization is not consistent in its relationship to grammaticalization (Diewald 2011; Heine 2013; Degand & Evers-Vermeul 2015). “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.” (Degand & Evers-Vermeul 2015: 59). This discussion therefore goes far beyond the descriptive scope of the present paper.

5.3 Further development as a verb

In §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 *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 §5.1, based on the noun ‘thing’ and a demonstrative, which serves as both a nominal and a verbal placeholder (Dimock 2010: 207–209).

- (46) *Dá-me aquela 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 Tupí-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 *maʔe* as a hesitator has also been noticed (Godoy

2020: 209). 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 *ba?e* 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, *ba?e* 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 *ba?e* 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 *ba?e* 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 §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

²⁷Constructions 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.

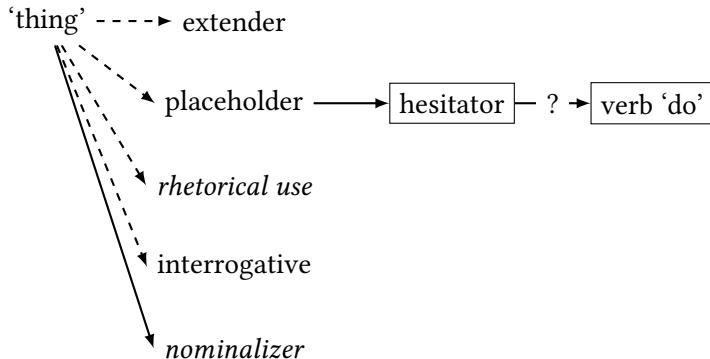


Figure 7: Diachronic developments from the noun ‘thing’ in Teko

that the noun for ‘thing’ has developed into an independent hesitator, very likely via the placeholder use of *ba?e*. Third, it has been suggested that the verb ‘do, make’ may in turn have derived from the hesitator.

As mentioned in §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

I would like to thank Kenza Van Heuzel, Romain Confrère & Viviane Ribes for their help in extracting data from Elan and segmenting it in Praat, Jennifer Krzownski for the Praat script and statistics, Matthew Stave, François Pellegrino and Rémi Anselme for technical advice, Meral Şeker for help with my English, two reviewers for their suggestions, and above all Brigitte Pakendorf for highly relevant comments and extremely stimulating discussions.

Abbreviations

COMPL	completive	EXH	exhortative
CONCES	concessive	EXIST	existential copula
CONT	continuative	FUT	future
CONTR	contrast	HES	hesitator
DEM	demonstrative	IDEO	ideophone
DIM	diminutive	IMP	imperative
EXCL	exclamative	INCL	inclusive

INDET	indeterminate	PL.O	plural of object
INJ	injunctive	POSTP	postposition
INTER	interrogative	PFV	perfective
INTENS	intensive	PURP	purpose
IRR	irrealis	RED	reduplication
LOC	locative	REF	referential
MAN.DEM	manner demonstrative	REL	relativizer
NEG	negation	RELN	relational
NSP	non-specific possessor	SG	singular
PAST	past	SOC.CAUS	sociative causative
PH	placeholder	SUB	subordinator
PL	plural	TRANSF	transfer
PL.S	plural of subject		

References

- Adler, Silvia & Estelle Moline. 2018. Les noms généraux : Présentation [General nouns: A presentation]. *Langue française* 198(2). 5–18. DOI: 10.3917/lf.198.0005.
- Allan, Keith & Kate Burridge. 1991. *Euphemism & dysphemism: Language used as shield and weapon*. New York: Oxford University Press.
- Association Solidarité Guyane. 2000. *Euro en Guyane. Langue amérindienne : Teko [The Euro in French Guiana. Amerindian language: Teko]*. Direction des relations avec les Publics et de la Communication du Ministère de l’Economie, des Finances et de l’Industrie.
- Boersma, Paul & David Weenink. 2023. *Praat: Doing phonetics by computer [Computer program]*. Version 6.3.11. <http://www.praat.org>.
- Burridge, Kate. 2006. Taboo, euphemism, and political correctness. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, Elsevier, 455–462. Cham: Elsevier.
- Cachine, Jean-Marc, Lucie Couchili, Didier Maurel, Eddy Monerville & Jammes Panapuy. 2020. *Kaleta tekō ’awu panaītsī ’awu ipope ma’ē. Dictionnaire tekō –français [Teko - Frenc dictionary]*. <https://corporan.huma-num.fr/Lexiques/dicoTeko.html>.
- Cheung, Lawrence Yam-Leung. 2015. Uttering the unutterable with *wh*-placeholders. *Journal of East Asian Linguistics* 24(3). 271–308. DOI: 10.1007/s10831-014-9130-x.
- Chousou-Polydouri, Natalia, Zachary O’Hagan & Lev Michael. N.d. *Mawetí-Guaraní lexical database*.

- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111. DOI: 10.1016/S0010-0277(02)00017-3.
- Cutting, Joan (ed.). 2007. *Vague language explored*. Basingstoke: Palgrave Macmillan.
- Cutting, Joan. 2019. German, Spanish and Mandarin speakers' metapragmatic awareness of vague language compared. *Journal of Pragmatics* 151. 128–140. DOI: 10.1016/j.pragma.2019.03.011.
- Daniel, Michael & Edith Moravcsik. 2013. The associative plural. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. DOI: 10.5281/zenodo.7385533.
- de Castro, Eduardo Viveiros. 1998. Cosmological deixis and Amerindian perspectivism. *The Journal of the Royal Anthropological Institute* 4(3). 469–488. DOI: 10.2307/3034157.
- Degand, Liesbeth & Jacqueline Evers-Vermeyl. 2015. Grammaticalization or pragmaticalization of discourse markers? More than a terminological issue. *Journal of Historical Pragmatics* 16(1). 59–85. DOI: 10.1075/jhp.16.1.03deg.
- Dienst, Stefan & David Fleck. 2009. Pet vocatives in Southwestern Amazonia. *Anthropological Linguistics* 51(3-4). 209–243. DOI: 10.1353/anl.2009.0001.
- Dietrich, Wolf. 1994. Word formation, syntax, or noun classification? Tupi-Guarani *mba'e* – between lexicon and grammar. *Revista latinoamericana de estudios etnolingüísticos* VIII. 113–128.
- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2). 365–390. DOI: 10.1515/ling.2011.011.
- Dimock, Laura. 2010. Fillers and placeholders in Nahavaq. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 119–138. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.06dim.
- ELAN. 2022. (*Version 6.3*) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>.
- Enfield, N. J. 2003. The definition of WHAT-d'you-call-it: Semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117. DOI: 10.1016/S0378-2166(02)00066-8.
- Fox, Barbara A. 2010. Introduction. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 1–9. Amsterdam: John Benjamins.
- Fronek, Josef. 1982. *Thing* as a function word. *Linguistics* 20(9). 633–654. DOI: 10.1515/ling.1982.20.9-10.633.

- Ganenkov, Dmitry, Yury Lander & Timur A. Maisak. 2010. From interrogatives to placeholders in Udi and Agul spontaneous narratives. In Nino Amiridze, Boyd H. Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 95–118. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.
- Gasparini, Noé. 2015. *Base de données lexicales et textuelles de siriono, Fieldwork Language Explorer [Siriono lexical and textual database, Flex]*. Unpublished.
- Godoy, Gustavo. 2020. *Os Ka'apor, os gestos e os sinais [The Ka'apor people, gestures and signs]*. Rio de Janeiro: Universidade Federal do Rio de Janeiro. (Doctoral dissertation).
- Grenand, Françoise. 1989. *Dictionnaire wayãpi-français [Wayãpi-French dictionary]* (Langues et Sociétés d'Amérique traditionnelle). Paris: Peeters / SELAF.
- Guérin, Valérie. 2015. Demonstrative verbs: A typology of verbal manner deixis. *Linguistic Typology* 19(2). 141–199. DOI: 10.1515/lngty-2015-0006.
- Halliday, M. A. K. & Ruqaiya Hasan. 2013. *Cohesion in English*. London: Routledge. DOI: 10.4324/9781315836010.
- Hayashi, Makoto & Kyung-Eun Yoon. 2006. A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. *Studies in Language* 30(3). 485–540. DOI: 10.1075/sl.30.3.02hay.
- Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6). 1205–1247. DOI: 10.1515/ling-2013-0048.
- Hennecke, Inga & Wiltrud Mihatsch. 2022. Between placeholder and filler: Degrees of prosodic prominence of French *machin* and *truc*. *Pragmatics & Cognition* 29(2). 297–323. DOI: 10.1075/pc.22003.hen.
- Jensen, Cheryl. 1998. Comparative Tupí-Guaraní morpho-syntax. In Desmond Derbyshire & Geoffrey Pullum (eds.), *Handbook of Amazonian languages*, vol. 4, 490–603. Berlin: Mouton de Gruyter.
- Jensen, Cheryl. 1999. Tupí-Guaraní. In R. M. W. Dixon & Alexandra Aikhenvald (eds.), *The Amazonian languages*, 125–163. Cambridge: Cambridge University Press.
- Kirjavainen, Minna, Ludivine Crible & Kate Beeching. 2022. Can filled pauses be represented as linguistic items? Investigating the effect of exposure on the perception and production of *um*. *Language and Speech* 65(2). 263–289. DOI: 10.1177/00238309211011201.
- Kleiber, Georges. 1987. Mais à quoi sert donc le mot *chose*? [But what is the word *chose* good for?] *Langue française* 73(1). 109–128. DOI: 10.3406/lfr.1987.6431.

- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. 2nd edn. Cambridge: Cambridge University Press. DOI: 10.1017/9781316479704.
- Luelsdorff, Philip A. 1995. A grammar of suspense. *Journal of Literary Semantics* 24(1). 1–20. DOI: 10.1515/jlse.1995.24.1.1.
- Maurel, Didier. 1998. *Eléments de grammaire émérillon [Elements of Emerillon grammar]*. Paris: Association d'ethnolinguistique amérindienne.
- Maurel, Didier. 2000. *Taila. 1596-1789 : Deux siècles de relations inter-communautaires teko-kaliña [Taila. 1596-1789: Two centuries of Teko-Kaliña relationships]*. Strasbourg: Université Marc Bloch, Strasbourg-II. (Master's Thesis).
- Maurel, Ti'iwan. 1991. *Dzawapinim o'olam tamandua [The leopard and the giant anteater]*. Red. by Gustave Ho Fong Choy (Krik Krak). Maripasoula: Association des Amis de l'Ecole élémentaire de Maripasoula.
- Mauri, Caterina & Andrea Sansò. 2017. Un approccio tipologico ai general extenders [A typological approach to general extenders]. In Marina Chini & Pierluigi Cuzzolin (eds.), *Tipologia, acquisizione, grammaticalizzazione. Typology, acquisition, grammaticalization studies*, 63–72. Milano: Franco Angeli.
- Mauri, Caterina & Andrea Sansò. 2018. Linguistic strategies for ad hoc categorization: Theoretical assessment and cross-linguistic variation. *Folia Linguistica* 52. 1–35. DOI: 10.1515/flih-2018-0001.
- Mihas, Elena I. 2019. Documenting ritual songs: Best practices for preserving the ambiguity of Alto Perené (Arawak) shamanic *pantsantsi* ‘singing’. *Language Documentation and Conservation* 13. 197–230.
- Mihatsch, Wiltrud. 2006a. *Machin, truc, chose : La naissance de marqueurs pragmatiques*. In Drescher, Martina & Frank-Job, Barbara (eds.), *Les marqueurs discursifs dans les langues romanes : Approches théoriques et méthodologiques*, 153–172. Bern: Peter Lang.
- Mihatsch, Wiltrud. 2006b. *Machin, truc, chose : La naissance de marqueurs pragmatiques [Machin, truc, chose: The birth of pragmatic markers]*. In Martina Drescher & Barbara Job (eds.), *Les marqueurs discursifs dans les langues romanes : Approches théoriques et méthodologiques*, 153–172. Frankfurt am Main: Peter Lang.
- Navet, Eric. 1994. Introduction. In Ti'iwan Couchili & Didier Maurel (eds.), *Contes des Indiens émérillon*, 1–11. Paris: Conseil International de la Langue Française.
- Overstreet, Maryann & George Yule. 2021. *General extenders: The forms and functions of a new linguistic category*. Cambridge: Cambridge University Press.
- Palacios Martínez, Ignacio M. & Paloma Núñez Pertejo. 2015. “Go up to miss thingy”. “He’s probably like a whatsit or something”. Placeholders in focus.

- The differences in use between teenagers and adults in spoken English. *Pragmatics* 25. 425–451. DOI: 10.1075/prag.25.3.05pal.
- Podlesskaya, Vera I. 2010. Parameters for typological variation of placeholders. In Nino Amiridze, Boyd H. Davis & Margaret Maclagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 11–32. Amsterdam: John Benjamins. DOI: 10.1075/tsl.93.02pod.
- Renault-Lescure, Odile, Françoise Grenand & Eric Navet. 1987. *Contes amérindiens de Guyane [Amerindian tales of French Guiana]* (Collection Fleuve et Flamme). Paris: Conseil International de la Langue Française.
- Rice, Alexander. 2025. *Mashti*: A multipurpose filler in Northern Pastaza Kichwa. In Brigitte Pakendorf & Françoise Rose (eds.), *Fillers: Hesitatives and placeholders*, 373–439. Berlin: Language Science Press. DOI: 10.5281/zenodo.15697595.
- Rodrigues, Aryon. 1984. Relações internas na família linguística Tupí-Guaraní [Tupí-Guaraní language family internal relationships]. *Revista de antropologia* 27-28. 33–53.
- Rose, Françoise. 2018. *Teko lexical and textual Toolbox database*.
- Rose, Françoise. 2011. *Grammaire de l'émérillon teko, une langue tupi-guarani de Guyane française [A Grammar of Emérillon Teko, a Tupi-Guarani language of French Guiana]* (Langues et Sociétés d'Amérique traditionnelle). Louvain: Peeters.
- Rose, Françoise. 2012. Borrowing of a Cariban number marker into three Tupi-Guarani languages. In Martine Vanhove, Thomas Stolz, Aina Urdze & Hitomi Otsuka (eds.), *Morphologies in contact*, 37–69. Berlin: Akademie Verlag.
- Rose, Françoise. 2021. Mojeño Trinitario. *Journal of the International Phonetic Association*. Illustration of the IPA 52. 562–580. DOI: 10.1017/S0025100320000365.
- Rose, Françoise. 2023. Word classes in Maweti-Guaraní languages. In Eva van Lier (ed.), *The Oxford handbook of word classes*, 729–751. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198852889.013.27.
- Rose, Françoise. 2024. Teko ideophones: Description of a word class. *Linguistic Typology at the Crossroads* 4(1). 91–146. DOI: 10.6092/ISSN.2785-0943/16117.
- Rose, Françoise & Sebastian Drude. Submitted. Maweti-Guaraní. In Patience Epps & Lev Michael (eds.), *Amazonian languages: An international handbook*. Berlin: De Gruyter.
- Santos-Granero, Fernando. 2007. Of fear and friendship: Amazonian sociality beyond kinship and affinity. *Journal of the Royal Anthropological Institute* 13(1). 1–18. DOI: 10.1111/j.1467-9655.2007.00410.x.
- Schegloff, Emanuel A. 1979. The relevance of repair to syntax-for-conversation. In Talmy Givón (ed.), *Discourse and syntax*, 261–286. New York: Academic Press.

- Seraku, Tohru. 2022. Interactional and rhetorical functions of placeholders: A relevance-theoretic approach. *Journal of Pragmatics* 187. 118–129. DOI: 10.1016/j.pragma.2021.10.023.
- Seraku, Tohru. 2024. Placeholders in crosslinguistic perspective: Abilities, preferences, and usage motives. *Linguistics* 62(6). 1473–1498. DOI: 10.1515/ling-2023-0068.
- Souza Mello, Antônio Augusto. 2000. *Estudo histórico da família linguística tupi-guarani. Aspectos fonológicos e lexicais [Historical study of the Tupi-Guarani language family. Phonological and lexical aspects]*. Universidade Federal de Santa Catarina. (Doctoral dissertation).
- Ti’iwan, Didier Maurel, Yves Blanc & Eric Navet. 1993. *Wilakala nūwā'e dzowoma 'ēkom omba'e*. Maripasoula: Association des amis de l’école élémentaire de Maripasoula.
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes* (Oxford linguistics 6). Oxford: Oxford University Press.
- Vallejos Yopán, Rosa. 2023. From demonstrative to filler: *Este* in Amazonian Spanish and beyond. *Linguistics* 61(3). 651–678.
- van der Auwera, Johan & Olga Krasnoukhova. 2021. How to do words with ‘things’: Multiple grammaticalization from ‘thing’ in Tupi-Guarani. *Italian Journal of Linguistics* 33(2). 69–98.
- Wojtylak, Katarzyna I. 2015. Fruits for animals: Hunting avoidance speech style in Murui (Witoto, Northwest Amazonia). *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 41(41). 545–562. DOI: 10.20354/B4414110007.
- Wojtylak, Katarzyna I. 2019. Talking to the spirits: A jungle-at-night register of the Murui people from the Northwest Amazon. *The Mouth: Critical Studies on Language, Culture and Society* 4. 78–90.
- Zikmundová, Veronika. 2013. *Spoken Sibe: Morphology of the inflected parts of speech*. Prague: Karolinum Press.

Name index

- Abondolo, Daniel, 134
Adelaar, K. Alexander, 171, 177, 203
Adelaar, Willem F. H., 394, 425
Adler, Silvia, 450
Aikhenvald, Alexandra Y., 6, 10, 14, 18, 99, 156, 287
Aikio, Ante, 134, 156
Albarracín de Alderetes, Lelia Inés, 425
Allan, Keith, 470
Allwood, Jens, 17, 82
Ameka, Felix, 430
Amiridze, Nino, 3, 5, 9, 275, 378
Anderbeck, Karl, 214
Anderson, Gregory D. S., 135
Anderson, Stephen R., 291
Aprilani, Herdian, 214
Aralova, Natalia, 44, 47, 50, 51, 53, 60, 75–78, 93, 105, 122
Ariel, Mira, 132, 161
Austin, Peter, 319

Baayen, R. Harald, 150
Bajo Pérez, Elena, 5
Baker, Brett J., 329
Barbera, Joseph, 398
Beattie, Geoffrey, 414
Bel'dy, Raisa A., 44, 52, 53, 63
Bendezú-Araujo, Raúl, 390
Beňuš, Štefan, 361
Betz, Simon, 417, 420

Billings, Blaine, 5, 7, 14, 16, 20, 22, 25–28, 80, 81, 89, 90, 104, 147, 180, 213, 214, 219, 220, 243, 244, 255, 261, 298
Blake, A. L., 5, 12, 13, 243
Bloomfield, Leonard, 2
Blust, Robert, 5, 177, 213
Boersma, Paul, 446
Boldyrev, Boris V., 45
Bordluk, Daisy, 318
Boula de Mareüil, Philippe, 316
Bourdieu, Pierre, 316
Braun, Angelika, 27, 104, 317
Breiman, Leo, 149
Bresnan, Joan, 319
Breunesse, Merlijn, 362
Brône, Geert, 410
Brown, Penelope, 132, 140
Brugmann, Karl, 362
Budrikis, Amy, 331
Bühler, Karl, 362
Bulatova, Nadežda Ja., 49
Bulatova, Nadezhda, 49
Bulgakova, Tatyana D., 44, 52, 53, 63
Burridge, Kate, 470, 471
Buskunbaeva, Liliya A., 243
Butterworth, Brian., 414

Caballero, Gabriela, 278
Cachine, Jean-Marc, 450, 453, 454, 464
Campbell, Lyle, 362

Name index

- Candea, Maria, 6
Carroll, Matthew J., 278
Chafe, Wallace L., 132, 141, 145, 368
Chelliah, Shobhana, 268
Chen, Victoria, 174
Cheung, Lawrence Yam-Leung, 5, 7, 8, 15, 55, 255, 297, 471
Chousou-Polydouri, Natalia, 444, 471
Cincius, Vera I., 51, 65, 78
Citko, Barbara, 138, 147
Clark, Herbert H., 2, 6, 16, 76, 116, 132, 141, 149, 155, 170, 196, 268, 281, 360, 378, 444, 457
Comrie, Bernard, 134, 136
Cordero, Luis, 425–427
Corley, Martin, 29
Corminboeuf, Gilles, 5
Couper-Kuhlen, Elizabeth, 132, 161
Crible, Ludivine, 2, 3, 16, 17, 105, 316, 330, 360
Crystal, David, 359
Cutfield, Sarah, 318
Cutting, Joan, 2, 450, 451
Däbritz, Chris Lasse, 5, 7, 10, 44
Daniel, Michael, 464
Davis, Joseph, 132, 141
de Castro, Eduardo Viveiros, 471
de Heer, Rolf, 321
de Jong, Nivja H., 316, 330
De Leeuw, Esther, 268
De Mulder, Walter, 279
Degand, Liesbeth, 475
Dienst, Stefan, 471
Diessel, Holger, 279, 362
Dietrich, Wolf, 442, 448
Diewald, Gabriele, 475
Dimock, Laura, 5–7, 9, 11, 12, 16, 26, 27, 121, 213, 472, 476
Dingemanse, Mark, 268, 269
Diver, William, 132, 141
Djigirr, Peter, 321
Döhler, Christian, 9, 11, 16, 22, 25, 28, 29, 83, 255, 277, 278, 280, 283, 286, 294, 295, 301, 302, 309, 416
Drude, Sebastian, 443
Du Bois, John W., 132, 141, 161, 172, 215
Duez, Danielle, 316
Egorova, Anastasia D., 44
Egorova, Maria A., 20
Egorova, Maria A., 267
Enfield, N. J., 5, 15, 16, 204, 226, 254, 255, 297, 302, 323, 378, 393, 470–472
Epps, Patience, 131, 146, 155, 160
Erker, Daniel, 2, 19
Evans, Nicholas, 286, 317, 318, 320, 325, 327, 329
Evers-Vermeul, Jacqueline, 475
Ewing, Michael C., 176, 213, 215
Fedorčenko, Alexander J., 134
Fehringer, Carol, 316, 330
Feyaerts, Kurt, 410
Fleck, David, 471
Forina, Michele, 150
Fortescue, Michael D., 134
Fox Tree, Jean E., 2, 6, 16, 76, 116, 132, 141, 149, 155, 281, 360, 378, 404, 444, 457
Fox, Barbara A., 2, 6, 132, 141, 170, 179, 230, 244, 374, 377, 378, 465
Fronek, Josef, 452

- Fry, Christina, 316, 330
- Gaby, Alice R., 4, 7, 16
- Ganenkov, Dmitry, 5, 7, 14, 18, 19, 41, 67, 68, 96, 116, 120, 121, 212, 213, 267, 472
- Garde, Murray, 323
- Gasparini, Noé, 445
- Gerhalter, Katharina, 19
- Gil, David, 216
- Givón, Talmy, 277
- Godoy, Gustavo, 445, 476
- Gogolev, Anatolij I., 134
- Goodwin, Charles, 410
- Goodwin, Marjorie Harness, 410
- Graziano, Maria, 29, 414
- Grebennjuk, Pavel S., 134
- Grenand, Françoise, 445
- Grenoble, Lenore A., 49, 134, 135
- Gries, Stefan Th., 150
- Grimes, Joseph E., 359
- Grzech, Karolina, 375, 390, 425
- Guérin, Valérie, 397, 428, 455
- Gullberg, Marianne, 29, 414
- Gusev, Valentin Yu., 44, 46–48, 62
- Hadar, Uri, 414
- Halliday, M. A. K., 359, 450
- Hanna, William, 398
- Harris, Alice C., 278, 362
- Hasan, Ruqaiya, 450
- Haspelmath, Martin, 244
- Hastie, Trevor, 152
- Haude, Katharina, 4
- Hayashi, Makoto, 3–8, 11, 14, 15, 18, 25, 29, 76, 77, 80, 118, 142, 147, 156, 169–171, 180, 186, 204, 211–214, 243, 244, 253, 275, 280, 283, 298, 316, 317, 341, 348, 349, 353, 354, 359, 360, 368, 369, 378, 386, 390, 395, 410, 444, 457, 460, 472
- Heath, Jeffrey, 4, 6, 9, 10
- Heine, Bernd, 267, 362, 427, 475
- Hennecke, Inga, 3, 5, 14, 17, 26–28, 81, 89, 196, 424, 452, 474, 475
- Himmelmann, Nikolaus P., 141, 174, 184, 229, 279, 280, 294, 302, 365, 366
- Hlavac, Jim, 45
- Hoetjes, Marieke, 414
- Hölzl, Andreas, 43, 50, 52, 54, 78
- Hömke, Paul, 412
- Hopper, Paul J., 132, 141, 362
- Huang, Huei-ju, 16
- Idiatov, Dmitry, 50
- Ikegami, Jiro, 44
- Jacques, Guillaume, 4
- Jehoul, Annelies, 16, 82
- Jensen, Cheryl, 442, 448, 453
- Jochelson, Waldemar, 135
- Johnsen, Laure Anne, 5
- Joseph, Brian D., 2
- Jurafsky, Dan, 5, 30, 89, 147, 149, 262, 265, 308
- Kalinina, Elena, 78
- Kantarovich, Jessica, 135
- Kany, Charles, 379
- Kapitonov, Ivan, 329
- Kasim, Yuslina, 214
- Katz, Aya, 156, 158, 161
- Kazakevich, Olga A., 42, 44–46, 48, 51, 53, 55, 56, 58–60, 135
- Kazama, Shinjiro, 44, 46, 48, 62, 63
- Keenan, Edward L., 291

Name index

- Keevallik, Leelo, 5, 15–17, 118, 316, 360, 378
Kendon, Adam, 82
Khasanova, Marina M., 44, 78, 105
Khurshudyan, Victoria G., 5, 7, 14, 26
Kibrik, Andrei Aleksandrovich, 262, 269
Kibrik, Andrej A., 5, 7, 30, 89, 147
Kim, Minju, 65
King, Gareth, 4
Kirjavainen, Minna, 2, 475
Kisa, Yağmur Deniz, 414
Kitano, Hiroaki, 4
Kitazawa, Shigeru, 412
Kjellmer, Göran, 361
Kleiber, Georges, 471
Kluge, Angela, 4, 7, 9, 14, 213, 261
Klyachko, Elena, 3, 5–7, 10, 14, 18–22, 30, 42, 44–46, 48, 50, 51, 53, 55–60, 76, 89, 92, 94, 97, 98, 105, 120, 122, 145, 243, 257, 263, 309, 310
Kogura, N. 児倉, 德和, 44
Kogura, Norikazu, 44
Korn, Agnes, 156
Korotaev, Nikolay A., 5, 7, 21, 27, 110
Kosmala, Loulou, 2, 3, 16, 17, 29, 82, 105, 316, 330, 360
Krasnoukhova, Olga, 441–443, 445, 448, 453, 454, 472–474
Krauss, Robert M., 414
Krejnović, Eruxim A., 136, 138
Kroll, Barbara, 359
Kuteva, Tania, 156, 267, 362, 472
Kuzmina, Raisa P., 56
Ladewig, Silva H., 420
Lahaussois, Aimée, 4
LaPolla, Randy J., 174
Larrivée, Pierre, 7
Laserna, Charlyn M., 267, 317
Lehmann, Christian, 428
LeSourd, Philip S., 5, 10, 111, 116, 118, 121, 142
Levelt, Willem J. M., 117
Lichtenberk, Frantisek, 286
Lickley, Robin J., 2, 75, 111, 112, 116, 120, 121, 374, 377, 405, 407, 412
Luelsdorff, Philip A., 467
Luk, Ellison, 318
Maclay, Howard, 2
Maisak, Timur A., 5, 18
Malchukov, Andrej L., 65
Mandel, Anne, 412
Maslova, Elena S., 133–138, 148, 157
Matić, Dejan, 5, 7, 14, 44, 50–52, 96, 102, 122, 135
Matras, Yaron, 30
Matthews, P H, 278
Maurel, Didier, 443, 445
Maurel, Ti'iwan, 445
Mauri, Caterina, 18, 59, 461, 463
Mayer, Thomas, 137–147, 158, 159
McDonnell, Bradley, 5, 7, 14, 16, 20, 22, 25–28, 80, 81, 89, 90, 104, 147, 171–176, 180, 213, 214, 217, 219, 220, 243, 244, 255, 261, 298
McKay, Graham, 329
Merlan, Francesca, 318, 325
Mihas, Elena I., 471
Mihatsch, Wiltrud, 3, 5–7, 14, 17, 26–28, 81, 89, 196, 424, 452, 474, 475
Mithun, Marianne, 5, 22, 25, 28, 132, 141, 317

- Moline, Estelle, 450
Moravcsik, Edith, 464
Mosel, Ulrike, 199, 215, 277
Mulac, Anthony, 132, 141
Muysken, Pieter, 375, 376, 394, 425
Myl'nikova, Klavdija M., 78

Næss, Åshild, 212
Nagasaki, Iku, 134, 136, 138–140, 156, 157
Nagaya, Naonori, 3, 5, 7, 14, 89, 120, 177, 204, 213, 243
Nakagawa, Hiroshi, 50
Nakano, Tamami, 412
Navarretta, Costanza, 5, 29, 122, 304, 405
Navet, Eric, 443
Németh, Zsuzsanna, 234
Nikolaeva, Irina, 44–46, 49, 54, 61, 65, 134, 136–147, 156, 158, 159
Noyce, Phillip, 321
Nuckolls, Janis B., 377, 390, 424, 439
Núñez Pertejo, Paloma, 5, 15, 27, 261, 262, 378, 444, 468, 471

Ono, Tsuyoshi, 132, 141
Orr, Carolyn, 375
Osgood, Charles E., 2
Oskolskaya, Sofia, 43, 78
Osorio, Sergio, 414
Ouattara, Aminata, 4, 6
Overstreet, Maryann, 18, 41, 59, 66, 461, 462, 464
Ozkan, Elif Ecem, 419
Ozoliņa, Larisa V., 45

Pakendorf, Brigitte, 5, 13, 16, 20–22, 25–30, 44, 47, 50, 51, 53, 60, 61, 64, 75–78, 93, 105, 122, 135, 147, 244, 261–263, 269, 316

Palacios Martínez, Ignacio M., 5, 15, 27, 261, 262, 378, 444, 468, 471
Panfilov, Vladimir Z., 50
Payne, Thomas E., 43, 52
Petrova, Taisija I., 44, 46
Pevnov, Aleksandr M., 44, 78, 105
Podlesskaya, Vera I., 3, 5–8, 14, 15, 21, 26, 27, 30, 42, 43, 49, 50, 61, 76, 77, 86, 89, 94, 104, 110, 120, 142, 144, 145, 147, 156, 157, 179, 191, 199, 211, 212, 230, 239, 244, 253, 262, 269, 308, 322, 324, 378, 391–393, 427, 444, 465, 472, 474, 475
Polivanov, Konstantin, 44, 46
Ponsonnet, Maïa, 3, 5, 7, 22, 25–27, 92, 104, 147, 261, 317–319, 321, 330, 406
Pyers, Jennie E., 414

Rauscher, Frances H., 414
Reeve, Mary-Elizabeth, 376
Renault-Lescure, Odile, 445
Rice, Alexander, 5, 7, 14, 20, 22, 23, 25, 28–30, 81, 82, 86, 101, 304, 375, 380, 439, 475
Rice, Sally, 430
Riesberg, Sonja, 174
Robbek, Maria E., 50
Robbek, Vasilij A., 50
Rodrigues, Aryon, 442, 443, 456
Rose, Françoise, 5, 7, 16, 18, 19, 22–25, 28, 59, 262, 269, 316, 406, 424, 442, 443, 445, 446, 451–453, 455, 464, 465, 468
Ross, Belinda Britt, 318

Name index

- Rossano, Federico, 410
Rubino, Carl, 5, 6, 13

Sakhatyla, 64
Salaoui, Abir, 19
Sankoff, Gillian, 132, 140
Sansò, Andrea, 18, 59, 461, 463
Santos-Granero, Fernando, 471
Saulwick, Adam, 325, 329
Savelyeva, Valentina N., 50
Schachter, Paul, 396
Schapper, Antoinette, 251
Schegloff, Emanuel A., 96, 111, 182, 185, 360, 466
Seifart, Frank, 377
Seraku, Tohru, 3, 4, 6–8, 14–18, 55, 255, 256, 323, 378, 466, 471
Seyfeddinipur, Mandana, 333, 417
Shakhovtsov, Kirill G., 44
Sharina, Sardana I., 56
Shattuck-Hufnagel, Stefanie, 121
Shopen, Timothy, 396
Sidnell, Jack, 380
Silverstein, Michael, 136, 316
Sjånes Rødvand, Linn Iren, 12
Skorik, Pjotr, 134
Smith, Linda Tuhiwai, 137
Šneider, Evgenij R., 44, 45
Souza Mello, Antônio Augusto, 442, 456
Stec, Kashmiri, 410
Stern, Nancy, 132, 141
Stewart, Oliver W., 29
Stivers, Tanya, 380
Stokhof, Wim A. L., 214
Stoyanova, Natalia M., 44, 53, 66
Streeck, Jürgen, 184, 229
Strobl, Carolin, 150
Suzuki, Ryoko, 410

Swanson, Tod Dillon, 377, 439
Sweetser, Eve, 410
Swerts, Marc, 361

Taksami, Chuner M., 50
Tanangkingsing, Michael, 16, 204, 243
Tárnyiková, Jarmila, 5, 15, 41
Thompson, Sandra A., 132, 141, 161, 410

Thornton, Warwick, 321
Ti’iwan, 445
Toldova, Svetlana Yu., 47, 48, 62
Tolskaya, Inna, 44, 49, 66
Tolskaya, Maria, 44–46, 49, 65, 66
Torero, Alfredo, 375
Traugott, Elizabeth Closs, 362, 471
Tree, Jean E. Fox, 170, 196, 268
Troiani, Giorgia, 141
Trousdale, Graeme, 471
Truong, Christina L., 175, 214, 217
Truppi, Chiara, 156
Trussel, Stephen, 5

Usher, Timothy, 251

Vajda, Edward J., 134
Vallejos Yopán, Rosa, 3, 5, 7, 14, 27, 90, 120, 121, 356, 378, 379, 390, 424, 457, 460, 475
van der Auwera, Johan, 441–443, 445, 448, 453, 454, 472–474
van Driem, George, 4
van Valin, Robert D. Jr, 136, 174
Vasilevich, Glafira M., 51
Vatanen, Anna, 136
Ventayol-Boada, Albert, 5, 8, 10, 19, 22, 23, 25–28, 30, 55, 104, 136, 142, 158, 257, 261

- Vidal-Covas, Lee-Ann Marie, 2, 19
Vinokurova, Antonina, 135
Visser, Eline, 7, 11, 16, 19, 22, 25–30,
 55, 104, 106, 145, 147, 253,
 257, 258, 260–262, 269, 298
Vogel, Petra M., 5, 10
Volodin, Aleksandr P., 134

Warren Carney, Lisa, 439
Watanabe, Honoré, 5, 14, 16
Weenink, David, 446
Wickham, Hadley, 380
Willerslev, Rane, 133
Williams, Nicholas, 19, 203, 213
Wojtylak, Katarzyna I, 471
Woodbury, Anthony C., 269
Woollams, Geoff, 219
Wouk, Fay, 5, 7, 10, 142, 182, 199, 203,
 213, 220, 243
Wrislej, Betsy, 375

Xelimskij, Evgenij, 134, 136

Yamada, Yoshiko, 44
Yoon, Kyung-Eun, 3–8, 11, 14, 15, 18,
 25, 76, 77, 80, 118, 142, 147,
 156, 169–171, 180, 186, 204,
 211–214, 243, 244, 253, 275,
 280, 283, 298, 316, 317, 341,
 348, 349, 353, 354, 359, 360,
 368, 369, 378, 386, 390, 395,
 444, 457, 460, 472
Young, Phil, 277
Yule, George, 18, 41, 461, 462, 464

Zhao, Yuan, 5, 30, 89, 147, 149, 262,
 265, 308
Zhornik, Daria O., 44, 46
Zikmundová, Veronika, 44, 54, 55,
 58, 470
Zorraquino, Martín, 379

Fillers

Fillers are non-silent linguistic devices used in disfluencies to gain time while searching for words. In addition, they are frequently used intentionally to avoid words for reasons of politeness, ‘conspirational’ motivations, or rhetorical purposes. Two syntactically distinct types of conventionalized fillers can be distinguished: placeholders and hesitatives (also called hesitators). Placeholders are referential and morphosyntactically integrated, while hesitatives are neither. Strikingly, even though fillers are cross-linguistically widespread, dedicated studies of such items in particular languages are still largely lacking.

This collective volume comprises in-depth descriptions of conventionalized fillers in a substantial variety of languages from Eurasia, Papunesia, Australia, and the Americas, hoping to stimulate typological research on fillers, both hesitatives and placeholders. The book aims to contribute to a better visibility of the topic among general linguists, to make data and analyses accessible that will be useful for further typological studies on the topic, and to provide models for descriptive linguists.

The introductory chapter discusses issues emerging from the previous literature and offers a new typology of fillers. It also highlights the major findings of the eleven remaining chapters. Each of these contains a detailed and typologically informed analysis of fillers in one or several underdescribed languages, based on corpora of natural speech and focusing on lexical fillers rather than on phenomena below the word-level (phonetic lengthening, truncation) or above the word-level (such as idioms and discourse markers like ‘you know’, or rhetorical questions like ‘what’s the word for that?’). The chapters cover a large amount of diversity, both in terms of languages and with respect to the type of filler. They focus on (i) the criteria for identification of the various types of fillers and the terminology used, keeping in mind that the domain is still largely under construction, (ii) a detailed analysis in terms of morphosyntactic distribution and, if possible, (iii) frequency in speech, and (iv) some reflection on the diachronic development of these disfluency markers.