Proposal for a collective publication on "Fillers: hesitatives and placeholders"
for the "Research on Comparative Grammar" series of Language Science Press

1. The topic

Fillers are non-silent linguistic devices used in disfluencies to gain time while searching for words (Fox 2010), such as "search sounds" like *um*, specific words like the Spanish demonstrative *este*, or discourse markers like *y'know*. Two syntactically distinct types of fillers can be distinguished: hesitatives (also called hesitators) and placeholders. Hesitatives are "paralinguistic signals" that do "not occupy any specific syntactic slot within the structure of an unfolding utterance", while placeholders are "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). The morphosyntactic integration of placeholders is illustrated by the following example taken from Podlesskaya (2010: 12).

(1) Russian

On	kupi-l	vsjakie	et-i …	pirožny-e.
he	buy-PAST.SG.M	various	PH-ACC.PL	cakes-ACC.PL

'He bought various PH [whatchamacallit] … cakes.'
Lit. 'He bought various them … cakes.'

Fillers are mainly seen as participating in self-repair (Schegloff, Jefferson, and Sacks 1977), and various fillers can indicate different disfluency functions (Clark and Fox Tree 2002; Kärkkäinen, Sorjonen, and Helasvuo 2007; Navarretta 2015) but they can also participate in interaction management and discourse planning (Hayashi & Yoon 2006: 527; Keevallik 2010: 164-165; Kosmala & Crible 2022). Furthermore, it has been argued that disfluency devices can be used intentionally with communicative goals, such as for generic expressions or as a vague identifier meaning 'N or whatever' (see for example Corley and Stewart 2008; Podlesskaya 2010).

Fillers, while very likely present in all languages, are infrequently described in grammars, especially for underdescribed languages. For instance, out of 237 tables of contents of grammars, "filler" is mentioned in five tables of contents, "hesitation" in two and "placeholder" in one, with zero mention of "hesitator" and "hesitative" (Taxogram tool at huma-num.fr, consulted June 06, 2023). As a consequence, the typology of the domain is only emerging. The main contribution to date is an edited volume (Amiridze et al. 2010) which includes five papers on individual languages -almost all Eurasian- and one paper offering a preliminary typology of placeholders (Podlesskaya, 2010).

The preliminary findings of Podlesskaya (2010) show that placeholders fill in for nouns more often than for verbs; in the latter case, verbal morphology can be attached directly to a pronominal root, to a derived stem, or to a bleached verbal root that combines with a pronominal or nominal stem. The morphosyntax of placeholders is often rather idiosyncratic, and languages vary in the degree to which the morphology of the target item is mirrored, from zero replication via only partial replication to the full complement of target morphology. They also vary in the degree of recycling (i.e. repetition) of preceding grammatical constituents such as prepositions.

Besides the synchronic morphosyntactic analysis of fillers, interesting questions are the relation of fillers to gestures (see for example Navarretta 2015), methodological issues in transcribing disfluencies in discourse, especially in the context of language documentation (Himmelmann 2006), and the historical development of such items. Mostly, placeholders develop out of (demonstrative) pronouns (like Russian *eto* in (1), see also [Hayashi and Yoon 2006]) or generic nouns (e.g. 'thing', as in Teko [Rose 2011: 176]), but lexicalized constructions that include an interrogative are also common, like English *whatchamacallit* (Podlesskaya, 2010; on the latter point see also Enfield 2003).

A particularity of this topic is that it lies at the crossroads of different approaches to language: cognitive studies of language production and perception, speech pathology (Davis and Maclagan 2010), conversational analysis (Ochs, Schegloff, and Thompson 1996), and grammar analysis. Within grammar analysis, it lies at the interface between syntax, semantics, and pragmatics (Enfield 2003). Probably because of this, it has often been considered a marginal phenomenon in linguistics (Ameka 1992). On the contrary, we consider that the analysis of fillers can be highly informative about the general organization of a language, for example in terms of constituency or parts of speech. They have been shown to be a useful heuristic tool for the analysis of word, morphology or discourse segments (Himmelmann 2014; H. Watanabe 2010; M. Watanabe 2002).

2. The contribution of the proposed publication

With this collective publication, our major aim is to instigate typological research on fillers, both hesitators and placeholders, on the basis of a larger variety of languages. We believe that publishing this research collectively will i) contribute to a better visibility of the topic among general linguists (since the initial volume edited by Amiridze et al. (2010), only a handful of papers have been published on placeholders [Cheung 2015, Vallejos 2021, Nagaya 2022, Seraku et al. 2022, Klyachko 2022] and hesitators [Kirjavainen et al. 2022, Kosmala & Crible 2022]), ii) make data and analyses accessible that will be useful for further typological studies on the topic, and iii) provide models for descriptive linguists.

All chapters will contain 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/discourse markers 'you know' or rhetoric questions 'what's the word for that?'). Stress will be put 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) a reflection on the diachronic development of these disfluency markers.

The proposed chapters cover a large amount of diversity, both in terms of languages and with respect to the type of filler. First, in terms of geographical distribution, the volume will cover four languages from Eurasia (all from Siberia), one language from Australia, three from Papunesia (two from Papua and one from Southeast Asia), one from North America, and two languages from South America. Second, in terms of the typology of disfluency markers, both dedicated placeholders and polyfunctional items are represented. The latter function not only as placeholders and hesitatives, but also as general extenders, a pro-verb, a euphemistic approximator and an attention-getter. Third, the following origins have been suggested for the fillers: the noun "thing", demonstrative and interrogative pronouns, the phrase 'so it is', and the question "what name"; however, in some languages the filler is a dedicated item with opaque etymology. The use of borrowed or code-switched fillers is also observed in oral corpora.

Individual papers furthermore address specific questions, such as how to distinguish the filler-use from the lexical use of the item in question (Döhler, Rose), the specific gestures that accompany different functions of the filler (Rice), borrowing of fillers (Kazakevich), variation in frequency and types of use between speakers (Pakendorf, Ponsonnet), and the developmental pathway of fillers (Mithun, Rose, Ventayol-Boada).

3. The editors

Brigitte Pakendorf and Françoise Rose are Senior Researchers of the CNRS in the Dynamique Du Langage (DDL) research unit in Lyon. Both are descriptive linguists, with a focus on Siberian and Amazonian languages, respectively, and have also regularly contributed to research in typology and areal/historical linguistics.

Françoise organized a workshop on fillers and placeholders at the DDL research unit in 2021 and 2022, consisting of 12 talks. She gave three talks within that workshop: one introductory talk on the topic, one talk on fillers in Teko and one on placeholders in Mojeño Trinitario. She also gave a conference paper on the Teko filler at the ALT meeting in Austin, TA, in December 2022, during a workshop on fillers and placeholders organized by Brigitte and Françoise. The workshop included 10 talks, eight of which will be represented in the volume proposed here (with three more chapters coming from participants in the DDL workshop or from colleagues who had to cancel their participation in the ALT workshop). Brigitte gave a talk on placeholders in Negidal in the local workshop organized by Françoise and a conference paper on the same topic at the ALT workshop.

The editors will write an introduction, presenting the topic, the state of the art, and summarizing the content of the chapters.

4. Contributors and abstracts

1. *Placeholders and hesitatives in Northern Selkup*, by Olga Kazakevich, Institute of Linguistics, Russian Academy of Sciences, Moscow

Northern Selkup placeholders and other fillers were mentioned in Selkup grammars, but no detailed description of this field has been done yet. The article presents an analysis of Northern Selkup fillers, both placeholders and hesitatives, mostly based on a text corpus recorded in Northern Selkup communities in 1996-2015. The most widespread placeholder in all the Northern Selkup dialects is *mɨ* 'thing, something' regarded in Selkup grammars as indefinite pronoun and presumably originated from the Proto-Samoyedic interrogative pronoun 'what'. This pronoun and its derivatives – adjectival *mɨ-ʎ*, adverbial *mɨ-ŋ* and verbal *mɨ-qo* – are used in a phrase substituting any noun, adjective, adverb or verb, whereby the morphological marking of this placeholder reflects the marking of the target words. *mɨ* can also be used as general extender. There are some other Selkup words and expressions often used as hesitatives: *na* 'this', *naɛnta* (*na ɛ:-nt-a* this be-EVID-3SG.SBJ) 'so it is', etc. In the function of hesitatives and even placeholders some borrowings or code-switched insertions from Russian are also used: *eto* 'this', *nu* ladno 'well', *yelki zelionyie* 'green fir trees'.

2. *Placeholders versus general extenders in Tungusic languages*, by Elena Klyachko, Higher School of Economics, Moscow

There are several placeholders in Tungusic languages: *aŋ(i/u), uŋ(un)* and the 'what' question word. The latter is also found in alternative questions (Tolskaya and Tolskaya 2008) and other "enumerations" as a general extender (Overstreet 2014). I will use oral corpora to compare mirroring in placeholders and general extenders (Podlesskaya 2010).
Placeholders mirror nominal inflectional categories: number, case and possession. They also copy frequent nominal derivational markers, e. g. caritive. As regards the verbal markers, person, number, and TAM can be copied but not the voice ((1) and (2)). However, general extenders usually copy all markers, including voice ((3)). Comparing data from Tungusic languages suggests a recent transformation of 'what' from a general extender to a placeholder.

(1) | *aja* | *ñaŋga=da* | *ñaŋga* | ***aŋi-nde-ze-m(i)*** | *sin-e-we* |
     |-------|------------|---------|----------------------|------------|
     | good | little=FOC | little | **PH-SEM-SUBJ-1SG** | 2SG-EP-ACC |

[kesu-li-nde-ze-m(i)]
torture-INC-SEM-SUBJ-1SG

‘Ok, **let me do this**, <u>let me torture</u> you a little’ (*Udihe*, [Nikolaeva and Tolskaya 2001])

(2) *ə-wki:* *minə* ***e:-ra*** *<u>əyi-fko:n-ə</u>*
     NEG-PHAB 1SG.ACC **what-PNEG** play-CAUS-PNEG
     ‘She did not **let me do this**, <u>let me play</u>’ (*Evenki*)

(3) *toɣo-wo* *o:kin=da:* *e:-wko:əru-t* *<u>jaŋu-w-rə</u>*
     fire-ACC when=FOC NEG-PTCP bad-ADVZ <u>talk-PASS-PNEG</u>
     ə-wko: ***e:-w-ra=da:***
     NEG-PTCP **what-PASS-PNEG=FOC**
     ‘One should not <u>talk</u> to fire badly or **do similar things**’ (*Evenki*)

3. *A corpus-based investigation of fillers in Negidal (Northern Tungusic)*, by Brigitte Pakendorf, Dynamique Du Langage, CNRS & Université de Lyon

Overt speech disfluencies include PLACEHOLDERS that substitute for a particular word and HESITATIVES, paralinguistic signals that overtly mark hesitation (Hayashi & Yoon 2006); both are here subsumed under the umbrella term FILLER. As shown by an investigation of a corpus of oral recordings, in the Northern Tungusic language Negidal (negi1245) the filler *uŋun* is frequently used: 885 tokens are found in ~78,000 words, amounting to ~11 tokens/1000 words—double the frequency of the filler found in corpora of Russian and Chinese (Podlesskaya 2010: 12). *Uŋun* primarily signals difficulties with speech production, rather than functioning in interaction management or self-repair. This can be deduced from the fact that it is the most frequent overt marker of disfluency, but it hardly ever occurs in the clause-initial position associated with pragmatic functions of fillers, it rarely elicits a "collaborative achievement of word search" (Hayashi & Yoon 2006: 499), and it is rarely used in self-repair. As a placeholder, *uŋun* substitutes for both nouns and verbs; it carries inflectional, but hardly any derivational morphology. In 40% of examples it is not followed by its target; when it is, very little material is recycled. In this paper, I will describe the functions of *uŋun* in detail and discuss them from a cross-linguistic perspective.

4. *The placeholder in Kolyma Yukaghir*, by Albert Ventayol-Boada, University of California, Santa Barbara

Kolyma Yukaghir, one of the two extant Yukaghir languages spoken in Northeastern Siberia, exhibits the placeholder *l'ə-*, which speakers use when a specific lexical item has momentarily eluded them. Typically, it targets nouns, verbs, and demonstratives, but it might also be able to substitute other lexical categories. In discourse, the placeholder is attested rather infrequently; in the 19 texts I analyzed from Nikolaeva and Mayer (2004), I retrieved

94 placeholder tokens of a total of 3,513 words (i.e., 2%). Of these, in 74 instances (or 80%) the relevant lexical item is later retrieved, and in most of these cases (55 or 74%) the placeholder mirrors all morphological categories that are later specified on the retrieved lexical item.

The origin of the placeholder, however, is unclear, but it does not resemble other sources identified (Hayashi & Yoon 2006, Podlesskaya 2010, Nagaya 2022). In fact, the placeholder coincides in form with one of the two copulas in the language (Krejnovich 1982; Maslova 2003; Nikolaeva 2006; Nagasaki 2010). This scenario begs the question of whether the resemblance is a historical accident of two independent forms converging into the same phonological shape, or whether they are synchronically distinct but historically related.

5. *Describing the placeholder in Dalabon, a polysynthetic language of Australia*, by Maïa Ponsonnet, CNRS/Université Lyon 2

This article analyzes the placeholder *keninjhbi* in Dalabon, a polysynthetic language of the Gunwinyguan family (non-Pama-Nyungan, northern Australia), based on 476 occurrences extracted from a 60-hour corpus.

Keninjhbi can replace verbal as well as nominal targets, and can receive practically any of the morphology available to it. It can mirror the target's marking either integrally or partially; some occurrences carry informative morphology that does *not* recur on the target.

The rich and flexible morphology of the Dalabon disfluency-management tool leaves speakers free to communicate more or less syntactic information in the absence of a lexical form. Reflecting this flexibility, 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 comparison between speakers' respective strategies suggests that in a polysynthetic language like Dalabon, the placeholder allows speakers to achieve distinctive disfluency-management styles.

6. *Form and frequency of Kalamang fillers*, by Eline Visser, University of Oslo

This contribution focuses on the form and frequency of fillers in Kalamang (Greater West Bomberai, Papuan), based on a 5200-word corpus. It reveals more forms for both placeholders and hesitators than previously described for Kalamang. The main placeholder is *neba* 'what', which can replace all word classes. It is used with a much higher ratio than previously calculated (13.5 per 1000 words instead of 6), which is also much higher than reported in other languages, such as Russian and Chinese (5 and 6.68 per 1000 words, respectively). The most common hesitator is *nain* 'like', followed by *a* 'ehm'. Coding of the phonetic form of all hesitatives revealed many different pronunciations, however, with clear preferences for [nain] or [nainː] and [aː], respectively. This data can inform choices in the

development of a standardized spelling. Finally, the data reveal big individual differences in frequency of use of fillers (ranging from 13 to 103 per 1000 words), stressing the need for speaker-balanced corpora.

7. *Hesitators and placeholders in Komnzo,* by Christian Döhler, Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin

This chapter addresses the topic of fillers in Komnzo, a Yam language spoken in Southern New Guinea. Fillers in Komnzo can be divided into hesitators (like English 'um', 'er') and placeholders (like English 'whatchmacallit', 'thingie'). After a brief description of hesitators, the chapter focusses on the placeholder pronominal *bäne* and its related forms. I show that *bäne* has developed from and is in some parts of its paradigm identical to the medial demonstrative *bäne*, but that it differs from it phonetically as well as morpho-syntactically.

Phonetically, it can be differentiated by a break in the intonation contour (*bäne kabe* 'that man' vs. *bäne … kabe* 'who's-that … the man'). Morpho-syntactically, the demonstrative *bäne*, as all demonstratives in Komnzo, is restricted to a subset of case flags, whereas the placeholder *bäne* (with its second root *baf*) can be inflected for all cases in their animate as well as inanimate forms. Hence, the placeholder *bäne/baf* behaves more like the personal pronouns in Komnzo. I argue that it should be analysed as a placeholder pronominal. The chapter is based on a 12-hour corpus of naturally occurring speech.

8. *Fillers in two Austronesian languages of Sumatra*, by Bradley McDonnell & Blaine Billings, University of Hawai'i

This paper describes fillers in two under-described Austronesian languages of southwest Sumatra, Nasal [nasa1239] and Besemah [cent2053], based on documentary corpora of everyday conversations in each language. While these languages share similar grammatical properties, the fillers that they employ differ. Nasal employs the medial demonstrative pronoun as both a hesitative and placeholder, while Besemah employs a dedicated root for hesitative and placeholder functions, which can be reconstructed to the Proto Malayo-Polynesian as both a filler and a question word meaning 'what'. In both languages, the fillers are well integrated into the morphosyntax of the language. They occur in both nominal and verbal positions in the clause and readily combine with all affixes. In conversation, these fillers are commonly used as delay devices in word searches where it is the site of repair. There are, however, many cases where the filler is used without any evidence of repair, including instances where lexical items are avoided. We show that despite using different fillers, Nasal and Besemah speakers employ them in similar ways. The primary aim of this paper is to fill in the gap of our knowledge of fillers by providing a thorough description of two Austronesian languages of insular Southeast Asia.

9. *Placeholders on the Move*, by Marianne Mithun, University of California at Santa Barbara

The life cycles of fillers have rarely been documented, largely because speakers do not consider them elements with the same status as nouns and verbs. Accordingly, they do not appear in the philological record of most languages and are rare in written documents and even transcriptions of spoken language. Hayashi and Yoon (2010) note, however, that placeholders, lexical items filling the slot of a delayed word or constituent, often develop out of demonstratives. Comparisons across related languages and internal reconstruction can sometimes reveal likely pathways of evolution. The demonstratives in the Iroquoian languages of North America are not all cognate, but all have developed via similar strategies, from adverbials plus a verb: 'here it is'. They can occur as appositives to a following nominal and as placeholders. They also frequently occur at the ends of intonation units as speakers formulate the next idea. This pattern has evolved into a conventionalized discourse construction used to manage the flow of information, where the demonstrative, with non-final prosody, signals that further elaboration is to come. Here it need not mirror the grammatical status of what follows. This construction could, furthermore, serve as a basis for the crystallization of complement and relative clause constructions.

10. *One more thing 'thing' can do in Tupi-Guarani languages : 'thing' as a filler in Teko*, by Françoise Rose, CNRS/Université Lyon 2

This paper intends to open an avenue for the investigation of **maʔe* 'thing' as a filler in the Tupi-Guarani languages. It offers a first detailed analysis of the discourse uses of the reflex of **maʔe* in a Tupi-Guarani language, namely the *baʔe* form in Teko (emer 1243). This study is based on a corpus of spontaneous texts (>5000 words) collected by the author in the field. In terms of discourse uses, *baʔe* is essentially used as a hesitator (without morphology) for nouns and less often for other parts-of-speech. Prosodically, its realization as a hesitator differs from those as a noun in the average durations of its final vowel, the preceding pause, and the next word. Three other discourse uses have not been mentioned before in the literature on Tupi-Guarani languages : i) as a general extender after a list, meaning 'etc', ii) as a first (euphemistic) approximation before a reference to a supernatural being, carrying the appropriate morphology for the delayed constituent, and iii) with a directive clitic, as an attention-getter. The theoretical goal of this paper is to map paths of pragmaticalization of the Tupi-Guarani generic noun **maʔe* towards disfluency markers and other types of discourse markers in order to complement the map of its grammaticalization paths (van der Auwera and Krasnoukhova 2021: 90).

11.   Mashti: *a multipurpose filler in Northern Pastaza Kichwa*, by 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 a nominal *shuti* 'name'.

In discourse, *mashti* can serve three functions: as a non-referential and distributionally unrestricted hesitative, as a nominal and verbal placeholder (which can fully mirror 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 gaze aversion, prosodic prolongation, 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.

5. Proposed timeline

- before late December 2023: submission of manuscripts to volume editors
- reviewing by 1 external reviewer + 1 volume editor
- revisions
- acceptance of papers
- submission of volume to LSP by end of July 2024
- check by series editors
- typesetting, proofreading
- publication by end of 2024