Phonological Exaptation and Epenthesis in Tetun Terik

Tamisha L. Tan (Harvard University, Nanyang Technological University)

1. Overview The lexicon of Tetun Terik (Central MP: East Timor) contains a notable typological oddity – a vanishingly small number of the language's verbs begin with a vowel (< 1% of all verbs), an asymmetry not found in its nominal or adjectival inventories. Closer investigation reveals that the descendants of Proto-Malayo-Polynesian *V-initial verbs have innovatively prothesised non-etymological initial /h/ in Tetun Terik (1a-b). Verbs that began in PMP **h/q/R*, consonants which where uniformly lost in initial position, were similarly remade (1c).

(1) a. **ala(q)* > ***h**ola* 'to get' (expected ***ala*), cf. Welaun *ola*, Manggarai *ala*, S. Mambae *a:l*
b. **inum* > ***h**emu* 'to drink' (expected ***imu(n)*), cf. Welaun *enu*, C. Mambae *eun*
c. **qudip* > ***h**ori-k* 'to dwell, reside', ***h**ori-s* 'to live, be alive' (expected ***ori*)

In addition, Tetun Terik innovated an epenthetic [h] at certain prefix boundaries, e.g. after causative **pa-* > *ha-* (2), in lieu of the more expected [ʔ] epenthesis elsewhere in the language.

(2) a. **ama* > *ama-n* 'father' → *ha-**h**aman* 'to treat as one's father'
b. **isi* > *isi-n* 'body, fruit, contents' → *ha-**h**isi* 'to put in, fill'

This paper proposes a diachronic treatment of the development of epenthetic [h] and prothetic /h/ in Tetun Terik based on i) Multiple Exponence and reinforcing morphology (Harris 2017, Caballero & Inkelas 2018), and ii) the pressure for a vowel hiatus repair strategy at morpheme boundaries, resulting in iii) phonological exaptation (Lass 1990, G. Ramat 1998, Kaźmierski 2015), where morphologically irregular material was reinterpreted as phonologically predictable.

2. Tetun Terik Verbs An empirical survey of the Tetun Terik verbal lexicon reveals only 8 V-initial verbs out of an inventory of ≈840 (Hull 2001; Williams-van Klinken 2019), all of which may be excluded on independent grounds from being truly native vowel-initial verbs, either due to being loans (e.g. *ukun* 'to govern' < Malay *hukum* 'law' and *uru* 'to deal cards' < Malay *urus* 'to manage'), marginal dialectal variants (e.g. *oho* 'to kill', var. *hoho* and *uhu* 'to pluck fruit', var. *huhu*), or predicatively-used nouns (e.g. *udan* '(to) rain', *isin* '(to) fruit'). Others like *uma* 'to kiss (directed to children)' are onomatopoeic, while *iha* 'to be at' is a locative existential preposition.

Crucially, none of these words may be the base of the agent noun deriving prefix *mak-*, showing that they do not meet its categorial selectional requirement of verbhood. Most revealingly, no Tetun Terik dialect inflects these V-initial words in for subject agreement, which is otherwise obligatory on C-initial verbs in the 1SG and all /h/-initial verbs. I argue that this synchronically unexpected distribution is due to the historical reanalysis of *V-initial verbs as /h/-initial.

3. Epenthetic [h] In Tetun Terik, all vowel hiatus outside of a word stem must be resolved. Thus, [ʔ] epenthesis occurs across compounds, word boundaries, reduplicants (3a) and proclitics (3b).

(3) a. ida∼ida-k [idaʔidak]
RED∼one-ADJ
'each'

b. Na= ina-n [naʔinan]
3SG= mother-GEN
'His mother' (van Klinken 1999: 47)

In contrast, at **affix** boundaries, Tetun Terik employs [h] epenthesis instead (2). I argue that this latter strategy of [h] epenthesis was innovated on the basis of phonological contraction of the causative prefix **ha-* in the prosodically weak pretonic initial position, with loss of the atonic prefix vowel > *h-*. This led to insufficient salience of the causative exponent, resulting in secondary reinforcement and reapplication of the full prefix as *ha-h-*, schematised in (4).

(4) PMP **pa-qudip* > Pre-Tetun **ha-uri* >$_{\text{contract.}}$ **h-uri* >$_{\text{reinf., backing}}$ *ha-hori-s* 'to give birth'

This resulted in the appearance of multiple exponence, where causativity seemed to be realised twice on a single base as both *ha-* and *h-*. Following work by Yang (2002, 2016) and Albright (2002) on evidence for phonological generalisation over morphological subregularities in acquisition, and the general simplicity bias in phonological rule learning (Pycha et al. 2003), I assume that speakers have a functional preference for learning predictable phonological processes over irregular morphological ones (e.g. base suppletion created by diachronic change). Especially given that this reinforcement resulted in redundancy, I propose that the no-longer-transparent *h-* prefix was reinterpreted as a regular, phonologically-conditioned epenthetic consonant inserted for hiatus repair. This constituted a form of phonological exaptation (Lass 1990, Kaźmierski 2015) or 'refunctionalisation' (Booij 2010), where a formerly morphological marker was renewed as a serving phonological function instead, as has been argued for the reinterpretation of Latin inchoative *-i/esc* suffix as a prosodic stress-alignment marker in Romance (G. Ramat 1998).

4. <u>Prothetic /h/</u> Once innovated, the [h] epenthesis strategy was extended to other prefixal hiatus boundaries in which such reinforcement was not expected, including with inchoative *na-*, negator *la* in *la-[**h**]os* 'indeed not', and derivational prefix *da-* in (5). This extension is fully predicted by its now-phonological distribution.

(5) a. **ina* > *ina-n* 'mother' → *da-**h**ina-n* 'childless married woman'
b. **ama* > *ama-n* 'father' → *da-**h**ama-n* 'childless married man'
c. **qulu* > *ulu-k* 'first, formerly' → *da-**h**ulu-k* 'first time, initial'

The high frequency of epenthetic [h] led to the wide-scale reanalysis of all *V-initial verbs as properly /h/ initial as illustrated in (6) below, resulting in the gap discussed in Sec. 2 above.

Prothetic /h/

(6)

	Simplex		Prefixed
*#p > h	#hV	:	#ha-hV
*#V (*v.*)	#V → **#hV**	:	#ha-hV
*#V (*n./adj.*)	#V	:	#ha-hV

This prothesis was fed by analogy with etymologically *#p > h-initial bases (e.g. **picik* > *hisik* 'to sprinkle' → ***ha-h**isik* 'to make splash out'), which when causativised produced string-identical initial *hah-* sequences as reinforced V-initial verbs (e.g. ***ha-h**oris* 'to give birth' in (4)).

Crucially, this widespread prothesis only occurred on verbs, an asymmetry which I argue falls out from the relative frequency of vowel hiatus contexts. Namely, nouns and adjectives were only prefixed a minority of instances – when causativised, taking the associative *ka-* prefix, or undergoing /Ca/-reduplication. This sort of derivational morphology is generally optional. In contrast, inflectional morphology such as agreement prefixation is typically obligatory in Tetun Terik (even in imperatives). Thus, verbs occurred prefixed in all instances, except in citation contexts. The nearly universal appearance of this epenthetic [h] with verbal bases resulted in the back-formation of all and only *V-initial verbs as /h/-initial, while nouns and adjectives did not occur in hiatus contexts frequently enough to trigger the secondary innovation of /h/ prothesis.

5. <u>Selected References</u> Albright, A. (2002). Islands of reliability for regular morphology. Caballero, G. & Inkelas, S. (2018). A construction-based approach to multiple exponence. Harris, A. C. (2017). Multiple Exponence. Hull, G. (2001). Standard Tetum-English Dictionary. Kaźmierski, K. (2015). Exaptation and phonological change. van Klinken, C. (1999). A Grammar of the Fehan Dialect of Tetun. Lass, R. (1990). How to do things with junk: Exaptation in language evolution. Pycha et al. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. Williams-van Klinken, C. (2019). Tetun-English Interactive Dictionary. Yang, C. D. (2002). Knowledge and learning in natural language.