Morphological analysis when morphemes are hard to detect: evidence from infixation and reduplication

Dave Kenneth Tayao Cayado, Kathleen Rastle
Department of Psychology, Royal Holloway, University of London

There is substantial psycholinguistic evidence that visual word recognition involves morphological analysis, whereby morphologically complex words like *teacher* are analysed in terms of their constituent morphemes (Rastle & Davis, 2008). Existing theoretical accounts have argued that morphological knowledge arises from readers' ability to discover orthographic sequences that contribute predictably to the meanings of words through their text experience (Korochkina & Rastle, 2025; Plaut & Gonnerman, 2000). Stems occur repeatedly in words that have similar meanings (e.g. *cleanly*, *cleanliness*, *cleaner*) and affixes alter the meanings of words in predictable ways (e.g. *teacher*, *builder*, *publisher*) – and these forms of systematicity have consequences for learning. However, these claims have not been tested in a wider variety of languages. In fact, most evidence comes from suffixes in European languages like English (Cayado & Rastle, 2025), which are typically realised as consistent orthographic forms (e.g. *-ed* in *walked*, *jumped*). Further, they typically attach to full stems or to stems with minor orthographic alterations. This raises a critical question: does morphological analysis remain robust when surface orthographic forms are inconsistent?

Using the morpheme interference paradigm in Tagalog, we address this question by examining two such cases in Tagalog: infixation and reduplication. Infixes disrupt the consistent realisation of a stem because they are inserted *within* the stem (e.g., *tawag* → *t-in-awag* 'called' or *t-um-awag* 'to call' vs. *nag-tawag* or *tawag-in*). Reduplication involves partial or full copying of a stem, and therefore, the same morpheme is realised in thousands of ways (e.g., *tawag* and *lundag* become ***ta****tawag* 'will call' and ***lu****lundag* 'will jump', where *ta-* and *lu-* are instances of the same reduplicant morpheme). Eighty native Tagalog speakers performed a lexical decision task on four types of visually presented Tagalog pseudowords: [1] INFIXATION, where infixes <*in*> and <*um*> were inserted into existing stems; [2] REDUPLICATION, involving partial reduplication (copying the initial consonant-vowel cluster of the stem) or full reduplication (copying the entire stem); [3] PREFIXATION, where prefixes *ka-* and *na-* were attached to existing stems; and [4] SUFFIXATION, where suffixes *-an* and *-in* were attached to existing stems. Figure 1 provides examples of morphologically structured pseudowords in each of these conditions and their orthographic controls.

Generalised linear mixed-effects models revealed significant morpheme interference effects on accuracy across all four conditions, indicating that skilled Tagalog readers were less accurate when processing morphologically structured pseudowords than orthographic controls (see Figure 2A). Linear mixed-effects models further showed significant effects on reaction times for PREFIXATION, INFIXATION, and SUFFIXATION, but not for REDUPLICATION (see Figure 2B). Follow-up analyses of REDUPLICATION showed that partial reduplication elicited significant effects on both accuracy and reaction times, whereas full reduplication produced effects on accuracy only (see Figure 3A and 3B).

These findings show skilled readers' acquired sensitivity to morphological structure, despite lacking consistent surface realisation. This prompts necessary revisions for existing statistical learning accounts of morphology. Specifically, we show that readers can acquire morphological knowledge, even when the typical surface realisation of a stem is disrupted (infixation) and when the same morpheme has no stable surface realisation (reduplication). Perhaps, frequently encountering broken stems due to infixing of <*in*> and <*um*> have helped readers develop a more flexible letter position coding for the initial consonant of the stem, which allows them to appreciate that *tawag* in *na-tawag* is the same as in *t-in-awag*. Further, readers appear to track the systematic copying of CV sequences in ***lu****lundag* 'will jump' ***ta****tawag* 'will call' and ***ti****tili* 'will scream' that signal 'incomplete action', allowing them to build representations of reduplicant morphemes. By examining an understudied language like Tagalog, we ultimately show that morphological analysis is much more robust than previously observed, motivating significant revisions to theories of morphological learning and processing.

Figure 1. Language background and experimental items. Morphologically structured pseudowords were created by combining existing affixes and stems that do not typically co-occur. Orthographic control pseudowords were formed by altering one letter of the affix. In the reduplication condition, partial reduplication involved copying the initial consonant–vowel cluster of stems that do not typically undergo reduplication and prefixing it to the stem, whereas full reduplication involved copying the entire stem.

	Morphologically Structured Pseudowords	Orthographic Control Pseudowords
Infixation	m**in**artes m**um**artes	m**un**artes m**im**artes
Reduplication	**ma**martes **martes**martes	**mu**martes **murtes**martes
Prefixation	**ka**martes **na**martes	**ki**martes **nu**martes
Suffixation	martes**in** martes**an**	martes**un** martes**en**

Figure 2. The morpheme interference effect on accuracy (A) and reaction times (B) in each condition.

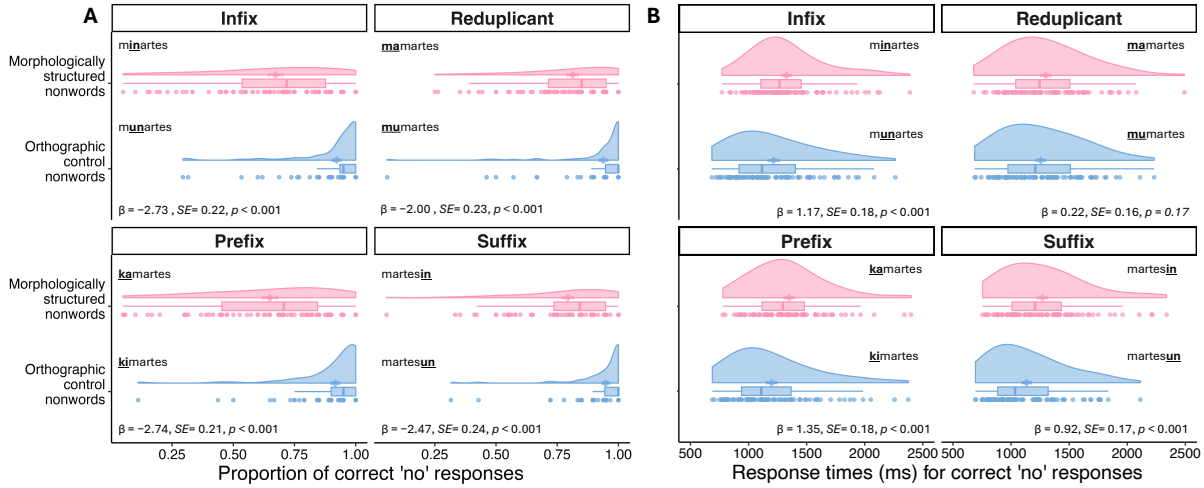

Figure 3. The morpheme interference effect on accuracy (A) and reaction times (B) in full and partial reduplication.