PHONOLOGICAL ANALYSIS

Phonological conditions on variable adjective and noun word order in Tagalog

Stephanie S. Shih
University of Southern California

Kie Zuraw
University of California, Los Angeles

Tagalog adjectives and nouns variably occur in two word orders, separated by an intermediary linker: adjective-linker-noun versus noun-linker-adjective. The linker has two phonologically conditioned surface forms, *-ng* and *na*. This article presents a large-scale corpus study of adjective/noun order variation in Tagalog, focusing in particular on phonological conditions. Results show that word-order variation in adjective/noun pairs optimizes for phonological structure, abiding by phonotactic, syllabic, and morphophonological well-formedness preferences that are also found elsewhere in Tagalog grammar. The results indicate that surface phonological information is accessible for word-order choice.*

Keywords: Tagalog, word order, variation, corpus, phonology, allomorphy, morphophonology

1. Introduction. The order of adjective and noun in Tagalog is traditionally described as freely variable. Both orders—prenominal adjective (1a; we call this *adjective* + *noun*) or postnominal adjective (1b; *noun* + *adjective*)—are claimed to be possible and semantically interchangeable, 'without any apparent difference in meaning' (Schachter & Otanes 1972:121).

(1) a. magandá-ng babáe (adjective + noun)
beautiful-LINK woman
'beautiful woman'

b. babáe-ng magandá (noun + adjective)
woman-LINK beautiful
'beautiful woman'

Despite the general claims of free order and semantic interchangeability, some have put forth hypotheses about the conditioning of order variation. Schachter and Otanes (1972; hereafter S&O) point to pragmatic and semantic factors, such as givenness or adjective semantic class, which, they observed, tend to prefer one order over another. Donohue (2007:360) instead suggests that phonological factors, rather than syntactic ones, condition the word-order choice.[1]

A linker particle intervenes between the noun and adjective. Regardless of word order, the linker follows the first word and has two phonologically conditioned allomorphs: *-ng* [-ŋ], which attaches to the first word, or *na* [na], which is orthographically a separate word. The form *-ng* occurs following vowel-final words (2a). In words that

* Acknowledgments to research assistants Seth Ronquillo, Arienne Filio, and Kimberly Nachor for data checking; Vera Gribanova and audiences at UCLA, UCSC, LSA 88, AFLA 21, and SCaMP 2016 for discussion; and an anonymous referee and two editors for commentary. Ivan Tam developed the software that created the 2006 Tagalog Web Corpus. Authors are listed by order of contribution: SSS designed and performed the corpus experiments and contributed syntax-phonology theory; KZ worked with Tagalog consultants and data and contributed on Tagalog morphophonology. Both authors analyzed the data, discussed results and implications, and wrote the manuscript at all stages; both take full responsibility for the content of the paper.

[1] We are not aware of any claims about how the syntax of the two orders might differ, or whether one is derived from the other. What will be crucial is that the syntax has a preferred order. This could be because the preferred order is basic, and the other order is derived by movement. Or it could be that both orders have the same (mirror-image) structure, but the syntax prefers one linear order.

end in an alveolar nasal [n] or glottal stop [ʔ], the [n] or [ʔ] is deleted, and *-ng* is used (2b). All other consonant-final words take *na* (2c).

(2)	a. babáe	→	babáe-**ng**	‘woman’
	b. karaníwan	→	karaníwa-**ng**	‘ordinary’
	c. itím	→	itím **na**	‘black’

S&O (p. 122) suggest that word-order variation interacts with this phonologically conditioned linker. Suppose that one of the words involved should take the *-ng* allomorph (*bago* ‘new’) and the other should take the *na* allomorph (*titser* ‘teacher’); then the preferred order, they claim, is the one in which the *-ng*-bearing item comes first (*bago-ng titser*), so that the linker allomorph is *-ng* rather than *na* (*titser na bago*). In other words, there is a preference for the order that yields the *-ng* linker. Such a preference would make word-order variation dependent on the morpheme’s SURFACE phonological form.

Crosslinguistic studies of word order and construction variation have yielded evidence that morphosyntactic variation of this type is conditioned by numerous factors, including psycholinguistic, sociolinguistic, semantic, and morphosyntactic pressures. In comparison, phonological factors are generally considered to be not as important in conditioning word-order alternations. Some previous work points to higher-level prosodic conditioning of syntactic choices (e.g. Zec & Inkelas 1990, Teeple 2008, Anttila et al. 2010) and of morphological processes such as linearization (e.g. Tucker & Henderson 2010). However, lower-level segmental, phonotactic conditioning is left largely untouched, on the broad assumption that details of phonological form should not affect larger (morpho)syntactic processes (e.g. Zwicky & Pullum 1986a).

Given the observations from the previous literature (e.g. S&O, Donohue 2007), Tagalog adjective/noun order variation appears to be a case in which surface phonological form affects morphosyntactic variation. In this article, we present the first systematic, large-scale, empirical study of adjective and noun word-order variation in Tagalog, using a corpus of written Tagalog web text, from Zuraw 2006. Our study contributes three main findings. First, phonological well-formedness conditions influence word-order variation in Tagalog, in addition to semantic and usage factors. This phonological optimization is more prominent for the adjective + noun order, which our results show is the favored ‘default’ order. Second, word-order variation is conditioned in part by the phonological surface form of the linker particle, as S&O surmise. Third, the phonological factors that matter to word-order variation in Tagalog are ones that are active elsewhere in the Tagalog grammar, in static lexical patterns and in morphophonological alternations. This suggests that the phonological factors most likely to influence between-word morphosyntactic variation are those that also have a role in the word-internal grammar of a language. Taken together, the results have implications for understanding the interaction between phonology and morphosyntax in variable linguistic phenomena.

The article is organized as follows. The data are introduced in §2, and we then present the phonological and nonphonological conditioning factors that are included in the study (§3). Results of the web corpus study are presented in §4, and §5 includes a brief comparison to a small speech corpus of Tagalog. At the end (§6), we discuss the ramifications of the results, focusing on predicting which phonological factors are most likely to affect morphosyntactic variation.

2. DATA. We used a corpus of Tagalog/Filipino web text from Zuraw 2006. These two language names are sometimes used interchangeably. *Tagalog* tends to refer to the language of the Tagalog people, and *Filipino* to the Tagalog-based national language of the Philippines. *Filipino* can also refer to Tagalog as spoken in Metro Manila and other multiethnic cities, or simply to a prestige register of Tagalog (Nolasco 2007). The corpus consists of text on the web collected in 2004, from a variety of genres—forums, blogs, newspapers, commercial sites, and more. We performed some additional cleaning on the corpus, including boiler-plate stripping and (near-)duplicate removal using tools from WaCky (Baroni et al. 2009), and additional measures to exclude text written in Cebuano. The resulting corpus contains 47,114,971 word tokens and 105,464 word types.

Tagalog orthography represents pronunciation well, with each letter corresponding to one sound in most cases. Letter values are close to the IPA (International Phonetic Association 1999), except that *ng* = [ŋ], *r* = [ɾ], and *y* = [j]. The glottal stop [ʔ] is not spelled; it occurs predictably before any vowel that is written word-initially (*itim* = [ʔitim]) or following another vowel (*babae* = [babaʔe]). There is also a phonemic [ʔ] that occurs word-finally, in contrast with Ø, but is still not spelled. When relevant we include a glottal stop in our examples, but we write it as a superscript to emphasize that it is not part of the spelling. Stress, although contrastive, is not marked in normal spelling; in this article we use acute accent marks for stress, which we determined using a print dictionary (English 1986).

Fortunately, it is easy to see from a word's spelling which linker allomorph it should take, how many syllables it has (one per vowel letter), how many segments long it is, and what sound it begins or ends with (except [ʔ], which is not spelled, but [ʔ]-final words lose their final consonant and take the *-ng* allomorph anyway, so it is still possible to tell which linker any word spelled with a final vowel letter should take).

To obtain a subcorpus of noun/adjective data, we first extracted all nouns and adjectives from the SEAsite online Tagalog-English dictionary (SEAsite 2001), using the part-of-speech tags there and some additional manual checking. We did not use any word that the dictionary listed as being both a noun and an adjective. We searched the web corpus for all possible noun-linker-adjective and adjective-linker-noun sequences. However, we automatically excluded tokens under the following circumstances: the second word itself ended in a potential *-ng* linker (if the second word itself bears a linker, then the sequence might not be a simple noun phrase); there was punctuation anywhere within the sequence; or the item contained any words ambiguous between being a target noun or adjective and a common nontarget item (list provided in Appendix A).

To check the validity of the items, we first selected for hand inspection 1,204 unique noun/adjective pairs (683 unique nouns and 110 unique adjectives). They were selected to include the nouns and adjectives occurring most frequently in the data, as well as the most frequent pairs. Three Tagalog-English bilinguals checked these pairs, identifying eleven words as problematic. Some were adverbs rather than adjectives, and others involved ambiguities caused by a linker. For example, *noong* could be *noó-ng* 'forehead-*linker*' or *noóng* 'when-*past*'. Any adjective/noun pairs containing any of these problematic words were then excluded from our total data set.

We were left with 149,689 adjective/noun-pair tokens, representing 14,591 types (including the hand-checked items found to be valid). The pairs included 1,708 different nouns and 587 different adjectives. For each of the 14,591 types, we compiled the number of tokens with adjective + noun order, and the number with noun + adjective order.

There are three potential sources of noise we could not control. First, *X*-linker-*Y* might not form a constituent as we want; instead, *Y* could be part of a complex modifier. An example from the current web (not in the corpus) is shown in 3, where the adjective *armado* could be part of a much longer modifier.

(3) a. táo-ng armado
'person-LINK armed'

b. **táo-ng** [**armado** sa kanilang ika-apat na henerasyon mobile na aparato]
'person-LINK [armed with their fourth LINK generation mobile LINK device]'

Second, we do not know which of our tokens are being used as predicates. S&O (pp. 121–22) describe special ordering conditions for 'unmarked' predicates—that is, those that do not include a personal pronoun, deictic pronoun, or personal name (namely, noun + adjective order is preferred when the adjective and noun 'provide equally new or equally important information').

Finally, we do not know for our corpus items if either of the words in a token is focused or represents new information.

To code items for various parameters of interest, we first looked up stress patterns and final glottal stops in a print dictionary (English 1986). We then coded each word's length in segments; each word's length in syllables (i.e. number of vowels); the length difference between the two words of the pair; the initial segment of each word; whether the words end with a nasal consonant, another consonant, or a vowel; whether either word begins with [na]; whether the adjective begins with the prefix [ma]; which allomorph of the linker each word would take if ordered first; and whether one of the words is a numeral or one of S&O's 'limiters' (pp. 121, 141), which include terms that refer to quantity (*marami* 'many', *bahagi* 'part of') or order (*huli* 'last').

2.1. NATURE OF EVIDENCE: WRITTEN CORPUS DATA. We briefly discuss here the use of written corpus data. There are typically two types of objections raised about using written data in linguistic investigations. These two objections have opposite expectations for how the written modality should behave, but both are rooted in the belief that written data cannot truly reflect 'natural language'. The first objection is that writing does not reflect on-line processing, because it can be carefully calculated: writers have time to reread and recast their language such that the final product is MORE well formed, phonologically and otherwise, than spoken or signed language, which is perceived as more spontaneous and 'natural'. The second objection is that language users do not access phonological form when writing, since the speech sounds are not actually vocalized: under this reasoning, written data should NOT demonstrate or care about phonological well-formedness patterns. We attempt to dispel these objections here, but a full exploration is beyond the scope of this article.

Written data has been shown via a large literature to be legitimately governed by principles of linguistic grammar and to reflect phonological patterns found in speech. In fact, as Biber (1988) points out, the potential differences, if any, between written and spoken data are much less prominent than differences across genres and contexts within each modality. The most informative differences are those that cut across modality (e.g. registers, formality, narrative context, abstraction). In certain situations, both speech and writing can be spontaneous and unrevised; in other situations, both speech and writing can be carefully calculated, recast, and revised (as with hesitations, pauses, and recasting in speech) (see also Nunberg 1990).

Writing is often claimed to be too slow to reflect on-line processing of language. This may have been plausible for handwritten works, but typing can close the time gap. For their psycholinguistic model of language production, Bock and Levelt (1994:950) assume a typical speech rate of about two to three words per second, following Maclay and Osgood (1959). Given that typing is about one word per second for average users and upward of two words per second for fast typists, typing nears the rate of spoken production. Moreover, typing—both spontaneous and dictated—demonstrates information-theoretic flow and linguistic effects similar to those of speech: for example, sensitivity to word frequencies, contextual predictability, and phonological and morphological properties (e.g.

Gentner et al. 1988, Zesiger et al. 1994, Weingarten et al. 2004, Cohen Priva 2010, Vinson et al. 2017).

Previous research reveals that phonological phenomena found in speech also occur in writing. English, for example, avoids adjacent sibilants (Menn & MacWhinney 1984, Zwicky 1987), which can result in the use of alternative syntactic constructions: for example, *the rooms of the house* versus *the hou**se's** rooms* (e.g. Hinrichs & Szmrecsányi 2007, Szmrecsányi & Hinrichs 2008, Grafmiller 2014, Shih et al. 2015; see also discussion in §3.2). Szmrecsányi and Hinrichs (2008:302–4) demonstrated that this phonological effect occurs reliably across spoken and written modalities. They note that the effect is weaker in written than in spoken data, but this also varies with genre: sibilant avoidance is stronger in written news reportage, which Szmrecsányi and Hinrichs attribute to the 'colloquialization of the written norm' to become more like speech. Grafmiller (2014:485) also reports adjacent-sibilant avoidance in spoken and written English: some written genres had stronger sibilant avoidance than the spoken data, and some had weaker. Grafmiller (2014) further finds no significant difference between spoken and written modalities in the effect of rhythm on English genitive construction choice.

In this article, we operate under the view that the best way to support linguistic hypotheses is to triangulate from multiple sources of data across modalities, registers, and other dimensions that may affect the observation of grammar. Tagalog is an underresourced language, and large-scale linguistic data in all mediums is not available; furthermore, to completely compare all possible modalities, genres, and contexts is outside the scope of a single article. Instead, we provide evidence from a detailed and to-date the largest corpus of written Tagalog, which offers data across several different registers, genres, and contexts. A sample of 284 files, approximately 0.2% of the total corpus, was hand-checked and found to be 33% news, 27% blogs and diaries, 8% discussion forums, 5% humor, 4% religious commentary, 3% official statements, 2% Bible verses, and 18% from other categories.

If anything, written Tagalog is a better 'surrogate for phonology' than, say, French (cf. Miller and colleagues' (1997:68) study on French, based on prescriptive orthography), because the Tagalog orthography represents more of the surface phonology—certain phonological alternations are represented, and pronunciation is more directly correlated with the orthography. For completeness, we also investigate in §5 the ONLY spoken Tagalog corpus available at the time of writing, though it is insufficiently large to make quantitative conclusions. We plan to look for and also invite future evidence from experimental and other paradigms.

3. CONDITIONING FACTORS. Though adjective/noun word order in Tagalog is generally characterized as freely variable, some hypotheses regarding the conditioning of order have been put forth in the previous literature. S&O, for example, focus largely on syntactic and semantic factors that prefer one order over another. Donohue (2007), in contrast, suggests that phonological factors (e.g. weight) matter. Donohue further notes that the less-frequent noun + adjective order is used in expressing particular contrasts on the adjectives.

In this section, we introduce the factors that we examined, with their motivations and individual descriptive statistics. Note that the results for individual factors reported in this section do not take into account the full complex interactive nature of the data, and thus should be taken as a descriptive introduction to the word-order effects. In §4 we present a full multivariate model that includes all of the factors and controls for potential confounds. (Factors that were investigated but turned out not to have reliable effects in the multivariate model are discussed briefly in §3.3.)

3.1. General predictors.

Basic default-order preference. With 90% of tokens showing adjective + noun order, this order can be taken as the 'default' order. In our statistical model, this overall preference will be captured by the intercept. Given the default ordering preference, the effect of other conditioning parameters discussed below adjust order preference from this baseline.

Linker preference. Another overall trend was noted by S&O (p. 122): in a given adjective/noun pair, if one word ends in a vowel, [ʔ], or [n] (thus taking *-ng*), and the other ends in a consonant other than [ʔ] or [n] (thus taking *na*), then it is 'slightly preferred' for the word that takes *-ng* to come first, so that the linker allomorph is *-ng*. From our corpus, we give two examples in 4 that follow this trend. For instance, example 4a occurs fifty-six times in the corpus, whereas its alternative ordering, 4b, occurs only five times (4c occurs twice, whereas 4d is unattested in the corpus). In general, we find that the likelihood of an order is increased when it features the *-ng* linker form.

(4) increased likelihood of order | decreased likelihood of order

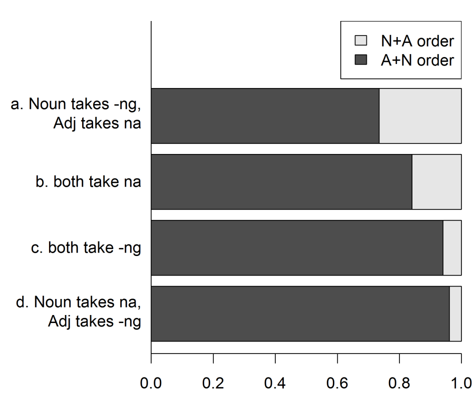
a. áso-**ng** ulól
dog-link mad
'mad dog'

b. ulól **na** áso
mad link dog

c. bágo-**ng** títser
new-link teacher
'new teacher'

d. títser **na** bágo
teacher link new

The plot in Figure 1 shows token counts broken down for each combination of potential linkers.[2] Cases like 4a,b are plotted as row (a), and cases like 4c,d are plotted as row (d). Adjective + noun order is always preferred in our data, but the preference is much weaker in row (a), where the default adjective + noun order produces the *na* linker. By contrast, the preference is stronger than usual in row (d), where the nondefault order produces the dispreferred *na* allomorph. The skewing between rows (a) and (d) is significant ($\chi^2(1) = 7161.71$, $p < 0.0001$).



Examples

	adj+noun	*noun+adj*
a.	ulól **na** áso	áso-ng ulól
	'mad dog'	
b.	hínog **na** bútil	bútil **na** hínog
	'ripe grain'	
c.	bágo-ng táo	táo-ng bágo
	'new person'	
d.	bágo-ng títser	títser **na** bágo
	'new teacher'	

Figure 1. Token counts by linker type.

Even though the preference for the *-ng* allomorph may not itself be phonologically driven—that is, it may be an arbitrary property of the linker—it does depend on phonology. To exercise a preference for *-ng*, it is necessary to know which linker allo-

[2] Here and in the following figures, the examples to the right of the plot correspond with each lettered plot bar.

morphs the adjective and noun would take if they came first, which in turn depends on what sounds they end with.

3.2. Phonological predictors. The following phonological predictors, ranging from phonotactic to morphophonological conditions, were considered: phonological weight (i.e. length), phonotactic markedness, morphophonological alignment, and syllable-structure optimization. These predictors were drawn widely from both the Tagalog literature and crosslinguistic studies of morphosyntactic and morphophonological variation. Effects found to be significant in the multivariate model (presented in §4) are discussed in depth here; effects not found to be significant are discussed briefly in §3.3.

Weight. Crosslinguistically, a common predictor of constituent ordering is the 'weight'—or length—of a constituent: it has long been observed that heavier constituents tend to occur at the peripheries of phrases (e.g. Behaghel 1909, Quirk et al. 1985, Hawkins 1994, Wasow 2002). English features the heavy-last principle, wherein heavier constituents tend to come later in phrases. In verb-final languages such as Japanese, by contrast, the weight effect is a heavy-first one (e.g. Hawkins 1994, Yamashita & Chang 2001).

We expect Tagalog to pattern with English, preferring heavy-last, because of its generally verb-initial word order. One explanation for the underlying cause of heavy-to-periphery phenomena is that moving heavier constituents to the edges aids cognitive processing. In right-branching languages like English or Tagalog, processing shorter constituents first minimizes the amount of material that a speaker or hearer has to keep in working memory (Hawkins 1994, Gibson 2000, Temperley 2006).[3] When it comes to relative clause order in Tagalog, which is also variable, S&O (p. 123) note a 'tendency to prefer the order head-linker-modifier when the modifying phrase is long'—that is, a heavy-last preference.

We operationalized heaviness as the number of segments (vowels or consonants) in a word. (We also looked at number of syllables, a highly correlated measure, and found similar results. We chose to use segments to allow for the possibility that consonants—both in onsets and in codas—can contribute to weight effects; for in-depth discussion of onset weight, see for example Topintzi 2006, Ryan 2011, 2014; for discussion of coda weight, see for example Hayes 1995 and references therein.) We expect that increasing the length of a noun (e.g. *kapangyaríhan* 'power', shown in 5a) will increase the likelihood of adjective + noun order. Likewise, we expect that increasing the length of an adjective (e.g. *pansamantalá* 'temporary') will increase the likelihood of noun + adjective order.

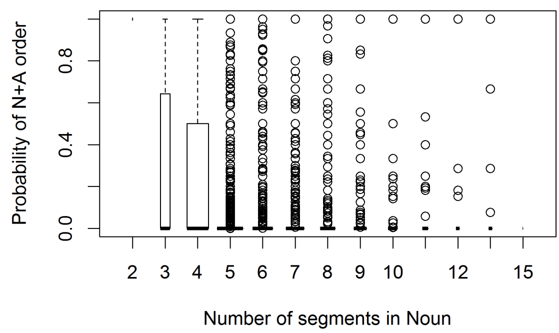
(5) a. dakíla-ng kapangyaríhan (expected to be more frequent than usual)
great-LINK power
'great power'

b. kapangyaríha-ng dakíla? (expected to be less frequent than usual)
power-LINK great
'great power'

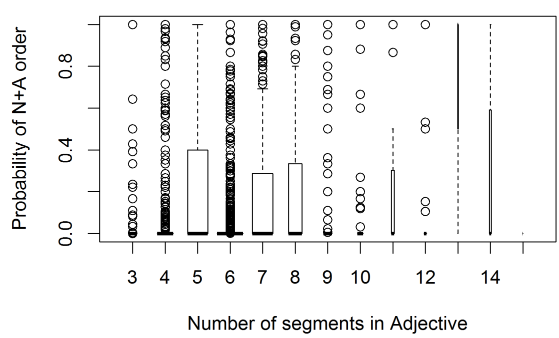
The boxplots in Figure 2 demonstrate the probabilities of noun + adjective order (labeled here 'N+A'), by the length of adjectives and nouns. The line at the center of each box indicates the median probability for that category, but these medians are always close to zero in this data, and thus are not distinguishable from the bottom of the box; this means that most pairs have a low rate of noun + adjective order in general. The top

[3] See for example, Szmrecsányi 2004, Shih & Grafmiller 2011 for discussion of competing explanations and measures of weight.

and bottom of each box are the twenty-fifth and seventy-fifth percentiles of probability for that category. The whiskers extend to the most extreme point that is no more than 1.5 times the height of the adjacent half of the box, and the circles show outliers beyond those bounds. The width of each box is scaled to the (square root of the) number of tokens that it depicts.



a. Probability of noun + adjective order (versus adjective + noun order) by number of segments in the noun.



b. Probability of noun + adjective order (versus adjective + noun order) by number of segments in the adjective.

FIGURE 2. Probabilities of noun and adjective order by weight.

What we find in the data is that the pattern for nouns is as the heavy-last principle predicts: longer nouns are significantly less likely to occur in first position (Wilcoxon $W = 1{,}102{,}856{,}871$, $p < 0.0001$), as shown in Fig. 2a. For adjectives, however, the data demonstrates a pattern opposite of what is expected: as adjective length increases, noun + adjective order is LESS likely and adjective + noun is MORE likely ($W = 765{,}499{,}276$, $p < 0.0001$). (Note: this effect is not as visually evident in Fig. 2b, but will become clearer in the multivariate model presented in §4.) As is discussed in §4, there is also an interaction, such that for long nouns, increasing the length of the adjective even further promotes adjective + noun order: this effect likely stems from a stronger preference for the default order as processing demands increase.

PHONOTACTICS: NASAL OCP. It is common for phonologies to obey the OBLIGATORY CONTOUR PRINCIPLE (henceforth, OCP), avoiding sequences of adjacent identical or similar elements (Goldsmith 1976, among many others). OCP effects have been seen before in interaction with syntax and morphology. For example, in the English genitive alternation, adjacent sibilants resulting from the combination of a sibilant-final possessor noun phrase and the possessive clitic are avoided (e.g. Menn & MacWhinney 1984, Zwicky 1987, Hinrichs & Szmrecsányi 2007, Shih et al. 2015): for example, *the rooms of the house* versus *the hou**se's** rooms* (as noted above). See also clitic movement in Warlmanpa to avoid nasal sequences (Wolf 2008:228).

In Tagalog adjective/noun ordering, if either of the words in a pair begins or ends with a nasal, there is a possibility of producing a sequence of nasals. For example, for the adjective/noun pair in 6, the order in 6b has a nasal-nasal sequence [m n], because the word placed first ends in a nasal and takes the *na* linker. Despite being in the default adjective + noun order, the likelihood of 6b is decreased when it exhibits this nasal-nasal sequence: 6b is less frequent in our data than 6a. Similarly, 7b, which is the less frequent order in our data, has a nasal-nasal sequence [ŋ n] (orthographically *ng n*), because the word placed first has the *-ng* linker and the second word begins with [n]. While we expect that noun + adjective orders are already less likely in comparison to the default adjective + noun orders, it is possible that the preference to avoid nasal sequences will further decrease the likelihood of those noun + adjective orders that exhibit nasal sequences.

(6) nasal-C + *na* could be penalized
  a. noun + adjective — b. adjective + noun
  pelúka-ng itim — iti**m n**a pelúka — 'black wig'

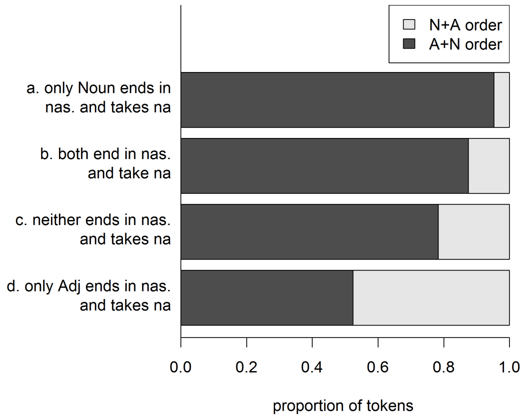
(7) *-ng* + nasal-C could be penalized
  a. adjective + noun — b. noun + adjective
  naturál na prodúkto — prodúkto-**ng n**aturál — 'natural product'

Looking at the phonotactics of Tagalog roots, we see that a sequence of two nasal consonants is avoided. In a corpus of 4,294 native, nonreduplicated disyllabic roots (English 1986), 1,257 have a medial consonant cluster. Of those, 659 (52%) have a nasal as the first consonant, and sixty-six (5%) have a nasal as the second consonant. We therefore expect thirty-three roots (0.52 × 0.05 × 1257) to have a nasal-nasal cluster, if consonants combined freely. Instead, there are only two such roots (*lingming* 'confused' and *pangnan* 'basket'). Thus, not only phonological typology but also Tagalog phonotactics provide a precedent for penalizing nasal-nasal sequences.

In our adjective/noun data, we find that a nasal consonant followed by *na*, as in 6b, is disfavored. As can be seen in Figure 3, noun + adjective order is much more common than is generally expected, given default order behavior, in row (d), when the adjective ends in a nasal and takes the linker *na* (i.e. it ends with [m] or [ŋ], not [n]); putting the adjective second avoids the nasal-nasal sequence. The nondefault noun + adjective order is also even less common than usual in the case of row (a), when the noun ends in a nasal and takes the linker *na*. The difference between rows (a) and (d) in Fig. 3 is highly significant ($\chi^2(1) = 89.03$, $p < 0.0001$).

The comparison in 7 is confounded by the general preference for the *-ng* linker and the prevalence and behavior of *ma-* adjectives (see §3.4), so it is not possible to look at these cases in isolation, but in our multivariate model, there is significant avoidance of orders like 7b.

Contextual markedness: *NC̥. Crosslinguistically, sequences of a nasal consonant followed by a voiceless obstruent (NC̥ clusters) can trigger a number of phonological repairs, such as deletion, assimilation, and fusion (Pater 1996, 2001). Within Tagalog words, the *NC̥ constraint (Hayes & Stivers 1996, Pater 1996, 2001) is active at prefix-stem boundaries, as in many related Austronesian languages (see Zuraw 2010 for more): a stem-initial voiceless obstruent usually fuses with a preceding nasal, as in 8a, but a voiced obstruent usually does not, as in 8b.



Examples

	adj+noun	*noun+adj*
a.	túnay na ámang	áma**ng n**a túnay
	‘real elder/father’	
b.	ití**m n**a páyong	páyo**ng n**a itím
	‘black umbrella’	
c.	ulól na áso	áso-ng ulól
	‘mad dog’	
d.	ití**m n**a pelúka	pelúka-ng itim
	‘black wig’	

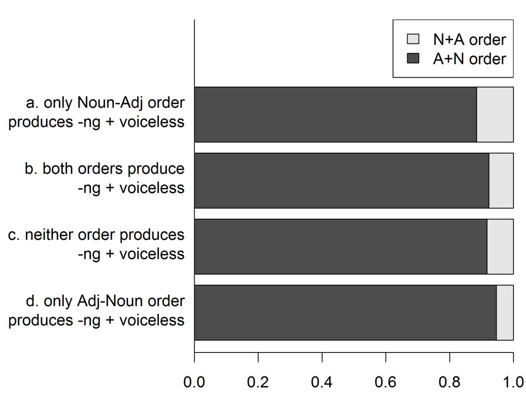
FIGURE 3. Avoidance of nasal C plus *na*.

(8) a. /pa**ŋ**-tabój/ → [pa**n**abój] ‘to goad’
 b. /pa**ŋ-d**ing/ → [pa**nd**inig] ‘sense of hearing’

The *NC̥ constraint is potentially relevant to adjective/noun ordering, because if the first word takes the *-ng* linker and the second begins with a voiceless obstruent, then an NC̥ sequence results, which could be avoided by the opposite word order.

(9) a. noun + adjective b. adjective + noun
 péra-ng nakalaán nakalaá-**ng p**éra ‘dedicated money’

Because *NC̥ is confounded with the linker allomorph, which we showed in §3.1 to have a strong effect on order, and various factors that depend on a word’s initial sound, we limit the data plotted here to cases in which both words take the linker allomorph *-ng* and begin with nonnasal consonants. The overall data shows the opposite of what we would expect: default adjective + noun order appears to be more common than usual when it produces an NC̥ cluster (Figure 4, row (d)). Our multivariate model will show instead that when other factors are controlled for, the *NC̥ effect is in the predicted direction.



Examples

	adj+noun	*noun+adj*
a.	kawáwa-ng bálo	bálo-**ng k**awáwa
	‘poor widow(er)’	
b.	túso-**ng p**láno	pláno-**ng t**úso
	‘shrewd plan’	
c.	bágo-ng líbingan	líbinga-ng bágo
	‘new cemetery’	
d.	nakalaá-**ng p**éra	péra-ng nakalaán
	‘dedicated money’	

FIGURE 4. Expected effect of *NC̥ does not hold in raw data.

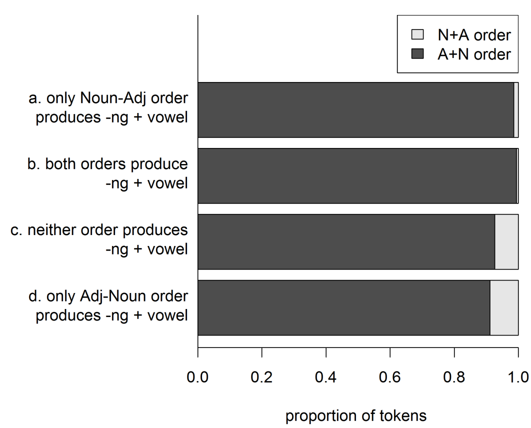
PHONOLOGICAL AND MORPHOLOGICAL ALIGNMENT. It is common for morpheme boundaries to prefer to coincide with syllable boundaries. For example, McCarthy and

Prince (1993) appeal to Align(Stem, Right; Syllable, Right), which requires the end of a stem to be the end of a syllable; they show alignment to be at work in Axininca Campa, Lardil, Hebrew, Bedouin Arabic, and Kamaiurá.

Alignment is relevant to Tagalog word order because if one word takes the linker *-ng* and the other begins with a vowel, as in 10b, then there are two attested pronunciations (S&O, p. 19). One inserts a glottal stop (e.g. [pu.lá**ŋ.ʔi**t.lóg]), in violation of Dep-C (Prince & Smolensky 1993).[4] The other has no glottal stop, resulting in marked syllabification. We know of no way to determine what the actual syllabification of this variant is, but either the word-final velar nasal must serve as an onset ([pu.lá.**ŋi**t.lóg]), violating Align, or we have a syllabification that violates both NoCoda and Onset ([pu.lá**ŋ.i**t.lóg]) (Prince & Smolensky 1993; see Tranel & Del Gobbo 2002 on *C.V). Putting the vowel-initial word first, as in 10a, will avoid the alignment problem, as long as the whole item is preceded by a vowel-final word or by a larger prosodic boundary.

(10) a. noun + adjective b. adjective + noun
itlóg na pulá pulá-**ng i**tlóg 'red [brined] egg'

In Figure 5, we see that noun + adjective order is even less common than expected, given the default preference for adjective + noun, when it produces *-ng* + vowel (row (a)). The difference between rows (a) and (d), where adjective + noun order produces the misaligned sequence, is highly significant ($\chi^2(1) = 714.86$, $p < 0.0001$).



Examples

	adj+noun	*noun+adj*
a.	ékstra-ng dalíri	dalíri-**ng é**kstra
	'extra finger'	
b.	ardmádo-**ng o**perasyón	operasyó-**ng a**rmádo
	'armed operation'	
c.	bérde-ng dáhon	dáho-ng bérde
	'green leaf'	
d.	pulá-**ng i**tlóg	itlóg na pulá
	'red egg'	

Figure 5. Morpheme and syllable alignment.

Syllable-structure optimization. In ordering coordinated noun pairs in English, two vowels in a row tend to be avoided (e.g. Wright et al. 2005, Benor & Levy 2006). In general, hiatus avoidance has widespread effects in phonology, and the constraint Onset is a fundamental one in optimality theory (Prince & Smolensky 1993:17).

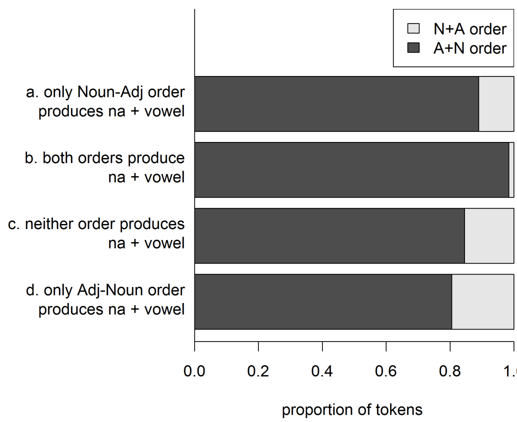
Hiatus avoidance is relevant to adjective/noun order when one word takes the *na* linker and the other begins with a vowel, as shown in 11. There are again two attested pronunciations: one violates Onset ([bá.gaj.na.**e**spesjál]), and one violates Dep-C (bá.gaj.na.**ʔ**espesjál).[5] The opposite order, 11a, can avoid the problem entirely.

(11) a. adjective + noun b. noun + adjective
espesyál na bágay bágay na $^{(?)}$**e**spesyál 'special thing'

[4] Or, if the glottal stop is underlying, in violation of the markedness constraint that drives the optional deletion, such as *PhraseMedialʔ.

[5] Or, again, *PhraseMedialʔ.

As the plot in Figure 6 shows, hiatus avoidance is a trend in our data. On the one hand, the default adjective + noun order is expected to be even more common than usual when the opposite order would result in misalignment, hiatus, or epenthesis (row (a)). On the other hand, the default adjective + noun order is expected to be avoided when this order would result in misalignment, hiatus, or epenthesis (row (d)). Comparison of these conditions show that the default order is indeed significantly more common in row (a) than in row (d) ($\chi^2(1) = 13.5$, $p = 0.0002$).



Examples

	adj+noun	*noun+adj*
a.	unát na bísig	bísig n**a u**nát
	‘outstretched arms’	
b.	espesyál n**a á**raw	áraw n**a e**speyál
	‘special day’	
c.	grábe-ng alítan	alíta-ng grábe
	‘serious disagreement’	
d.	pángit n**a á**sal	ásal na pángit
	‘ugly habit’	

FIGURE 6. Hiatus avoidance.

3.3. OTHER PHONOLOGICAL PREDICTORS. In addition to the factors above, we investigated other segmental and prosodic factors that were found not to have significant and reliable effects in the multivariate model, presented in §4. We summarize these briefly here.

ADDITIONAL SEGMENTAL FACTORS. Above, we showed that nasal-nasal sequences are avoided. The *-ng* allomorph of the linker is also velar and can thus produce a velar-velar sequence, as in *bálo-ng kawáwa* ‘widow-LINK poor’. We found that word order was not deployed to avoid this. In §6, we claim that this is because the phonotactics of Tagalog roots do not support a penalty for velar-velar sequences. This factor is included in our multivariate model below, but did not prove significant.

We also investigated the possibility that successive nasal onsets are avoided (e.g. *diláw* ***na manggá*** versus *manggá-ng diláw* ‘yellow mango’), since nonadjacent, similar consonants have been demonstrated to interact in other languages (e.g. harmony, dissimilation) (e.g. via McCarthy 1979, Hansson 2001, Rose & Walker 2004, Bennett 2013). Tagalog phonotactics support penalizing consecutive nasal onsets. There are 198 roots whose first onset is a nasal, and 565 whose second is a nasal, so we would expect eighty-nine roots whose onsets are both nasal, but there are only twenty-six. However, most nasal-initial adjectives begin with the prefix *ma-*, which has its own ordering preference (see §3.4). When this factor is included in the multivariate model (§4.2), controlling for the *ma-* confound, no significant effect is found.

ADDITIONAL PROSODIC FACTOR. A factor that we investigated but were not able to include in our multivariate investigation, due to Tagalog-specific limitations explained below, is stress rhythm. In English, a sequence of adjacent unstressed syllables (i.e. lapse) can be avoided through word-order choice (e.g. McDonald et al. 1993, Wright et al. 2005, Benor & Levy 2006, Shih 2014, Shih et al. 2015).

We might expect that Tagalog, a stress language, would have similar prosodic conditions on adjective/noun order. In Tagalog, primary stress is contrastive, mainly final or penultimate. If one word ends with a stressed syllable and the other begins with a stressed syllable, and they appear with the linker *-ng*, stress clash would result (12b, clash in bold). If one word ends with an unstressed syllable and the other begins with an unstressed syllable, stress lapse could result, and the unstressed linker *na* would further contribute to stress lapse (12d, lapse in bold).

(12) a. noun + adjective
σ́ σ σ σ́
báya-ng sawíʔ

b. adjective + noun
σ **σ́** **σ́** σ
sa**wí-ng bá**yan 'unfortunate country/people'

c. adjective + noun
σ σ σ́ σ σ́ σ
espesyál na áraw

d. noun + adjective
σ́ **σ** **σ** **σ** **σ** σ́
á**raw na espe**syál 'special day'

Secondary stress is not fully understood in Tagalog (e.g. French 1988, 1991), and it is usually not included in dictionary entries, so we did not annotate our data for it. We preliminarily examine disyllabic and trisyllabic words only, to avoid possible noise from unknown secondary stresses (further excluding any words whose stress is uncertain, including antepenultimate stress, which can be unstable).

The sharpest comparison comes from items where one order produces a long lapse (i.e. four unstressed syllables in a row) and the other does not produce a long lapse. Those data are plotted in Figure 7, and there is a difference in the expected direction. The rate of noun + adjective order is lowest when that order would produce a long lapse (row (a)), and the rate of adjective + noun order is lowest when that order would produce a long lapse (row (c)). The difference between these two cases is highly significant ($\chi^2(1) = 293.85$, $p < 0.0001$). (There are no cases where both orders produce a long lapse.)

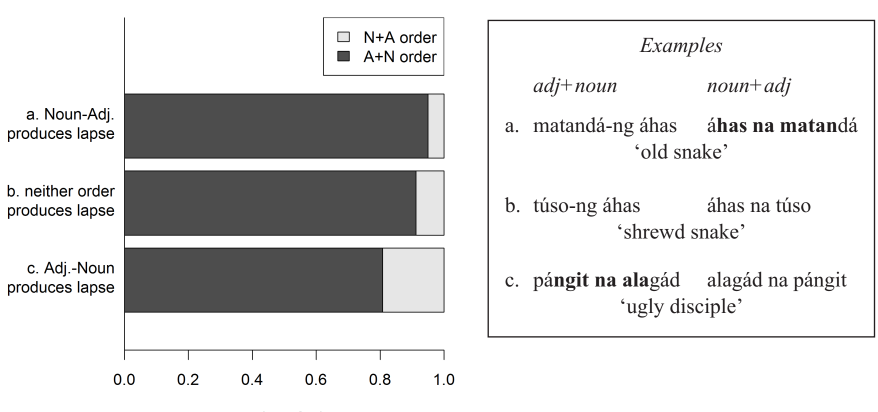


FIGURE 7. Stress lapse.

We similarly examined stress clash, but found that sequences of stressed syllables were NOT avoided and were actually more common than expected. We ultimately did not include rhythmic factors in the multivariate analysis presented in §4 because the nature of stress and word-internal rhythmic preferences is not fully understood for Tagalog. We reserve rhythmic factors and their potential relationship to word-order variation in Tagalog for future investigation, which will need to include a general investigation of rhythmic preferences throughout Tagalog phonology.

3.4. Nonphonological predictors. In addition to phonological factors, we coded for three nonphonological predictors that have been suggested from the previous literature on Tagalog word-order variation: adjective semantic class (i.e. quantifier versus nonquantifier), adjective prototypicality, and frequency. These predictors are discussed in turn below.

Quantifier adjectives. S&O note that adjective + noun order usually occurs when the adjective is an ordinal or cardinal numeral, or when it is part of a set of 'limiter' adjectives that 'express quantity, distribution, or order in a series' (p. 141). An example of a variable quantifier adjective from our corpus is given in 13.

(13)	adjective + noun	noun + adjective	
	lahát na táo	táo-ng **lahát**	'all people'

Depending on one's syntactic analysis, these words might not be classified as adjectives at all, and indeed they behave rather differently from other adjectives.

We coded adjectives as either quantifiers (numerals; and S&O's limiters, given in Appendix B) or nonquantifiers. Following S&O, the prediction is that quantifiers prefer to occur first even more strongly than other adjectives, and this was true in our corpus. Quantifier adjectives occur first 98% of the time; nonquantifiers occur first 89% of the time ($\chi^2(1) = 1659.09$, $p < 0.0001$).

Adjective prototypicality. Items that are prototypical are conceptually central to a given category: for example, for American English speakers, an apple is a more prototypical fruit than a loquat or a lemon. Prototypical items are more easily accessed in language processing, which results in earlier ordering: for example, *an apple and a loquat* is more likely than *a loquat and an apple* (Kelly et al. 1986, Onishi et al. 2008).

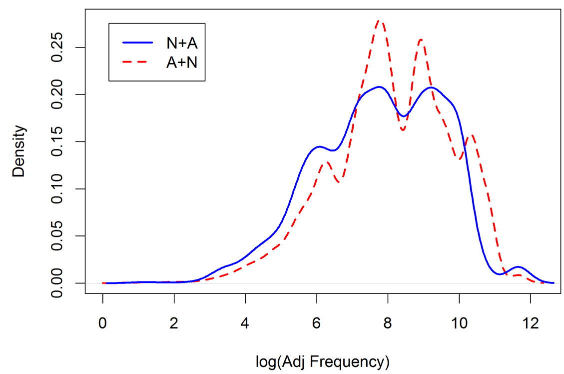
In Tagalog, adjectives are very often formed using the *ma-* prefix: for example, *mabaho* 'malodorous' from *baho* 'bad smell'. Thus, *ma*-initial adjectives are more prototypical as adjectives: an initial *ma* sequence signals adjective-hood. Given the potential increase in accessibility for prototypical items, we therefore expect to see that *ma*-initial adjectives (including ones that happen to begin with the phonological sequence *ma* but do not actually have the affix *ma-*: e.g. *mahal* 'dear') prefer the default adjective + noun order more strongly than adjectives that do not begin with *ma-*. In 45% of our tokens, the adjective begins with the string *ma*. Comparing only nonquantifier, *ma*-initial adjectives to other consonant-initial, nonquantifier adjectives, the difference is very small (87.9% *ma*-initial adjectives occur first; 87.3% non-*ma*-initial adjectives do), but significant ($\chi^2(1) = 9.33$, $p = 0.002$). The multivariate model below shows that when other factors are controlled for, the effect is significant.

Frequency. In processing, more-frequent items are hypothesized to be more readily accessible (e.g. Griffin & Bock 1998 and references therein); thus, more-frequent items should occur earlier. We control for frequency in our study by including individual adjective and individual noun frequencies, as well as adjective/noun-pair frequencies.[6]

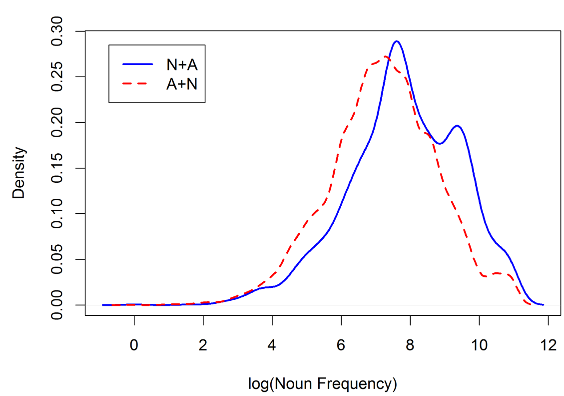
Our expectation is that the more frequent word will prefer to occur first: that is, default order should be even more frequent when the adjective is more frequent than the noun, and conversely, nondefault order will be more frequent than usual (even if still comparatively infrequent, given the default baseline) when the noun is more frequent than the adjective. The effects of individual adjective and individual noun frequencies are illustrated in the frequency density plots in Figure 8. In Fig. 8a, pairs of adjectives and nouns

[6] We suspect that there are more subtle effects of frequency that were not uncovered here; however, because frequency is not the focus of this article, we leave further investigation of its behavior for future work.

that consistently appear in the adjective + noun order (dashed line, *n* = 11,219) are shifted slightly higher on the adjective frequency scale, as compared to pairs that consistently appear in noun + adjective order (solid line, *n* = 2,498). This is what we expect: more-frequent adjectives have greater tendency to be first. Similarly in Fig. 8b, pairs with consistent noun + adjective order have more-frequent nouns.



a. Densities for adjective log frequencies for consistently noun + adjective pairs versus consistently adjective + noun pairs.



b. Densities for noun log frequencies for consistently noun + adjective pairs versus consistently adjective + noun pairs.

FIGURE 8. Effects of frequency of adjective and noun.

The plot in Figure 9 compares PAIR frequencies, again for pairs that appear consistently in one order. While very little effect is noticeable in the plot, the multivariate regression model (§4) finds that, after residualizing by individual noun and adjective frequencies, higher pair frequency promotes noun + adjective order (when the adjective would take *-ng*). We expect that pair frequency will correlate with the adjective and noun together being stored in a particular order, though we have no a priori expectation of whether that would be noun + adjective or adjective + noun. The regression model finds that it is noun + adjective for our data.

3.5. SUMMARY OF PREDICTORS. The predictors we tested using multivariate regression in the following section are summarized in Table 1.

Predictors that we did not examine, for practical reasons, but that we expect could have an effect on adjective/noun order variation include the following: being part of a

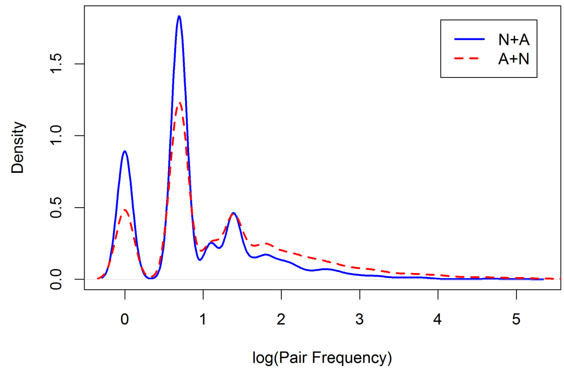


FIGURE 9. Densities for adjective and noun pair log frequencies for consistently noun + adjective pairs versus consistently adjective + noun pairs.

GENERAL CONDITIONING PREDICTORS	
Default order preference:	Adjective + noun preferred
Linker:	Adjective (*-ng* preferred)
	Noun (*-ng* preferred)
	-ng preferred over *na*
PHONOLOGICAL PREDICTORS	
Weight:	Adjective length
	Noun length
	Total length
*[nasal][nasal] OCP:	Avoid nasal-final adjective + *na*/*-ng*.
	Avoid nasal-final noun + *na*/*-ng*.
	Avoid *-ng* + nasal-initial adjective.
	Avoid *-ng* + nasal-initial noun.
*[velar][velar] OCP:	Avoid *-ng* + velar-initial adjective.
	Avoid *-ng* + velar-initial noun.
Long-distance nasal OCP:	Avoid *na* + nasal-initial adjective.
	Avoid *na* + nasal-initial noun.
Contextual markedness:	Avoid *-ng* + C̥-initial adjective.
	Avoid *-ng* + C̥-initial noun.
Alignment:	Avoid *-ng* + V-initial adjective.
	Avoid *-ng* + V-initial noun.
Hiatus avoidance:	Avoid *na* + V-initial adjective.
	Avoid *na* + V-initial noun.
NONPHONOLOGICAL PREDICTORS	
Quantifier adjective	Quantifier adjective goes first.
Prototypicality	Adjective that begins with *ma* goes first.
Frequency—Noun	Frequent noun goes first.
—Adjective	Frequent adjective goes first.
—Pair	Frequent pair prefers noun + adjective.

TABLE 1. Predictors to be tested.

predicate, givenness, focus, animacy, and stress rhythm. As mentioned in §2, an adjective/noun phrase used as a predicate has different ordering properties, according to S&O. If either the noun or the adjective represents discourse-given information, and the other represents new information, we also expect that word order could be used to make the new information more prominent (Quirk et al. 1985, Biber et al. 1999; though cf. Hinrichs & Szmrecsányi 2007, Shih et al. 2015). Similarly, if one of the words is under focus and the other is not, order could be affected. Finally, animacy has been found to be important in word order (e.g. Rosenbach 2005). The influence of these factors is saved for future investigation.

4. Results. We present here the results of a multivariate model of adjective/noun word order in the Tagalog web corpus. We used mixed-effects logistic regression modeling to examine the effects and reliability of each conditioning factor introduced in §3. Modeling was done using the glmer() function of the lme4 R package (Bates et al. 2013).[7] Model statistics were obtained using the MuMIn R package (Bartoń 2013); plots in the current section were generated using lattice and effects R packages (Sarkar 2008, Fox 2003, respectively). Binary predictors were centered by subtracting the mean, and numerical predictors were centered and standardized by dividing by twice the standard deviation. Centering and standardization help mitigate some potential multicollinearity and facilitate model convergence and direct effect-size comparisons between binary and numerical predictors (Gelman 2008). In the models, two random intercept terms—for nouns and for adjectives—were included to control for individual lexical items' propensity for one or the other order. A random intercept for each unique adjective/noun pair was also tested, but due to data sparseness for many pairs, once individual nouns and adjectives were accounted for, this pair term was dropped from the models.[8]

The final model was fit using stepwise backward elimination: insignificant factors were removed sequentially from the full model containing all of the factors. The statistical criterion for removal was if the coefficient was less than the absolute value of one and a half times the standard error (z-score $\in [-1.5, 1.5]$). If the factor was a part of a significant interaction term, but did not itself exhibit a reliable main effect, we did not remove it. We erred on the side of inclusion so as to be able to examine trends and between-factor interdependencies that were theoretically likely; furthermore, we leave trends in place so that they may inform future work in investigating factors important to Tagalog word-order choices.

The dependent variable was the log odds of nondefault noun + adjective order (e.g. positive regression coefficient indicates increased probability of noun + adjective order; negative regression coefficient indicates decreased probability of noun + adjective). In the logistic model, predicted probability of noun + adjective order is estimated as a linear combination of the intercept and factor weights, as in 14.

(14) $p(y) = \frac{1}{1 + e^{-(x' \cdot \boldsymbol{\beta})}}$,
where y = noun + adjective order;
$\boldsymbol{\beta}$ = vector of fitted coefficients (i.e. weights); and
$\boldsymbol{x'}$ = transposed vector of predictor variables, including 1 for the intercept.

The results of the final multivariate model are given in Table 2, where 'N' = noun, and 'A' = adjective.

The final model accounts for about 80% of the variance in word order (conditional R^2 = 0.7998). Much of this is accounted for by the random intercepts for individual

[7] The model reported here was estimated using lme4 version 0.999999-0 in R build 2.15.02. We tested the same model selections with lme4 version 1.1-7 in R build 3.1.1. There were no differences in reliability of predictors and only slight differences in effect size and model log-likelihood; however, model convergence is a known issue with the later lme4 versions, and the model reported here is the one that did not produce a false convergence warning.

[8] Models tested with a random intercept term for noun and adjective pair consistently produced false convergences.

FACTOR	EST	*SE*	*z*-VALUE	Pr (>|*z*|)	
GENERAL PREDICTORS					
i. intercept (order preference)	−2.4398	0.1661	−14.687	< 0.0001	***
ii. linker (N) = *na*	−0.7744	0.1449	−5.341	< 0.0001	***
iii. linker (A) = *na*	1.2793	0.2335	5.479	< 0.0001	***
PHONOLOGICAL PREDICTORS					
(nasal-initial A)	0.2240	0.3864	0.580	0.5621	
(C̥-initial A)	−0.1785	0.3127	−0.571	0.5681	
(C̥-initial N)	0.1795	0.1405	1.277	0.2016	
iv. segment length (N)	−0.1798	0.0405	−4.444	< 0.0001	***
v. segment length (A)	−0.3824	0.0956	−4.001	< 0.0001	***
vi. *[nas][nas]: nasal-final A	1.2240	0.3389	3.612	0.0003	***
vii. alignment, *hiatus: V-initial A	−2.3082	0.5619	−4.108	< 0.0001	***
GENERAL & PHONOLOGICAL INTERACTIONS					
viii. linker (N) = *na* * linker (A) = *na*	−0.2121	0.0671	−3.159	0.0016	**
ix. seg length (N) * seg length (A)	−0.0534	0.0079	−6.783	< 0.0001	***
x. *[nas][nas]: linker (N) = *na* * nas-initial A	0.2314	0.0931	2.484	0.0129	*
xi. *NC̥: linker (N) = *na* * C̥-initial A	0.1992	0.1014	1.966	0.0493	*
xii. *NC̥: linker (A) = *na* * C̥-initial N	−0.3299	0.0620	−5.317	< 0.0001	***
xiii. *hiatus: linker (N) = *na* * V-initial A	−0.3347	0.1873	−1.787	0.0739	.
NONPHONOLOGICAL PREDICTORS					
xiv. quantifier = Y	−3.2127	1.0457	−3.072	0.0021	**
xv. prototypicality: *ma*-initial = Y	−1.0521	0.2543	−4.138	< 0.0001	***
xvi. log(N frequency)	0.3467	0.0571	6.075	< 0.0001	***
xvii. A/N pair frequency (resid)	0.3941	0.1619	2.434	0.0149	*
NONPHONOLOGICAL INTERACTIONS					
xviii. linker (A) = *na* * A/N pair freq (resid)	−0.6540	0.2594	−2.521	0.0117	*

RANDOM EFFECTS	VARIANCE	*SD*
Noun (intercept)	4.5813	2.1404
Adjective (intercept)	6.1983	2.4896
N = 14,591	AIC_c = 31075.2	
Nouns = 1,708	Marginal R^2 = 0.1440; Conditional R^2 = 0.7998	
Adjectives = 587	κ = 4.6794	

*** significant at $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.1$

TABLE 2. Modeling estimates.

nouns and adjectives. Comparatively, the fixed effects alone account for 14% of the variance (marginal $R^2 = 0.144$). A lower fixed-effects R^2 is expected because, as discussed in §3.5, many semantic, syntactic, and sociolinguistic factors that could potentially influence Tagalog word-order variation have yet to be explored. The model does not exhibit harmful multicollinearity (κ = 4.6794).

Tables 3–5 summarize the hypothesized conditioning factors that were tested and indicate which factors were reliable predictors of adjective/noun order in our data set. Note that the regression results in Table 2 show the independent variables as they were coded for modeling; in Tables 3–5 and in subsequent discussion, the conditioning predictors we tested and found reliable are indexed to their corresponding factors in the model in Table 2 with small Roman numerals.

		INDEX	EFFECT?
Default order preference (adjective + noun)		(i)	✓
Linker:	Adjective (*-ng* preferred)	(iii)	✓
	Noun (*-ng* preferred)	(ii)	✓
	-ng preferred over *na*	(viii)	✓

TABLE 3. General conditioning predictors tested.

		INDEX	EFFECT?
Weight:	Adjective length	(v)	✓
	Noun length	(iv)	✓
	Total length	(ix)	✓
*[nasal][nasal] OCP:	Avoid nasal-final adjective + *na*/-*ng*.	(vi)	✓
	Avoid nasal-final noun + *na*/-*ng*.		✗
	Avoid -*ng* + nasal-initial adjective.	(x)	✓
	Avoid -*ng* + nasal-initial noun.		✗
*[velar][velar] OCP:	Avoid -*ng* + velar-initial adjective.		✗
	Avoid -*ng* + velar-initial noun.		✗
Long-distance OCP:	Avoid *na* + nasal-initial adjective.		✗
	Avoid *na* + nasal-initial noun.		✗
Contextual markedness:	Avoid -*ng* + C̥-initial adjective.	(xi)	✓
	Avoid -*ng* + C̥-initial noun.	(xii)	✓
Alignment:	Avoid -*ng* + V-initial adjective.	(vii)	✓
	Avoid -*ng* + V-initial noun.		✗
Hiatus avoidance:	Avoid *na* + V-initial adjective.	(xiii)	✓
	Avoid *na* + V-initial noun.		✗

TABLE 4. Phonological predictors tested.

		INDEX	EFFECT?
Quantifier adjective		(xiv)	✓
Prototypicality		(xv)	✓
Frequency:	Noun	(xvi)	✓
	Adjective		✗
	Pair	(xvii)	✓

TABLE 5. Nonphonological predictors tested.

Our results reveal that both phonological and nonphonological predictors play a role in conditioning variable adjective/noun order in Tagalog. Crucially, it is evident that not every effect that was tested (e.g. long-distance OCP) is a reliable predictor. We take this to be an indication that the reliable results that were found in this study are genuine data patterns: that is, it is not the case that ANY conditioning factor can elicit a result from the data modeling.

The model results in Table 2 show the direction, size, and reliability (i.e. significance) of the conditioning factors for predicting Tagalog adjective/noun ordering. We also tested the explanatory importance of each reliable conditioning factor using a drop-one, nested-models test of increase in second-order Akaike information criterion (AIC_c, 15; e.g. Burnham & Anderson 2002, 2004).

(15) $AIC_c = -2log(\mathcal{L}(\hat{\beta}|D)) + 2K + \frac{2K(K+1)}{n-K-1}$,
where $\mathcal{L}(\hat{\beta}|D)$ = maximum likelihood of observed data D given fitted parameters $\hat{\beta}$,
K = number of estimable parameters (i.e. constraints) in the model, and
n = sample size.

AIC_c quantifies how much information a model captures, given the evidence provided. It is calculated using model maximum log likelihood, and it crucially takes into account the number of parameters used in model estimation, rewarding parsimonious models that capture information with fewer parameters. Algebraically lower AIC_c values indicate that the given model captures more information, because AIC_c is inversely correlated with increasing model likelihoods. By dropping each predictor in turn, it is possible to examine how much a predictor contributes in the explanatory power of the full model.

An increase in AIC_c denotes that there is less evidentiary support—that is, more information not captured—for a model without the predictor in question than the full model.

The results of the drop-one AIC_c test are given in Figure 10. Following Burnham and Anderson (2004:271), '*' refers to at least a two-point increase in AIC_c value (ΔAIC_c), indicating that the predictor contributes to the explanatory power of the model, given the evidence provided. '**' refers to a three- to nine-point ΔAIC_c over the full model, indicating considerably more explanatory power contributed by the predictor removed. '***' refers to at least a ten-point ΔAIC_c from the full model: this indicates that when the tested predictor is removed, the resulting model receives no support as the best model of word order—that is, too much explanatory power is lost—when compared to the full model.

The AIC_c results provide a ranking of predictor importance. We see that phonological factors rank highly in their individual contributions to the model's explanatory power—in some cases, this is higher than nonphonological predictors. For example, the predictor for nasal-final adjectives (Table 2:vi, *[nas][nas] OCP) is by far the strongest; if it is removed, the resulting model would be far worse than a model from which the predictor for quantifier adjectives (Table 2:xiv) was removed. Individual predictor results are discussed in the following sections.

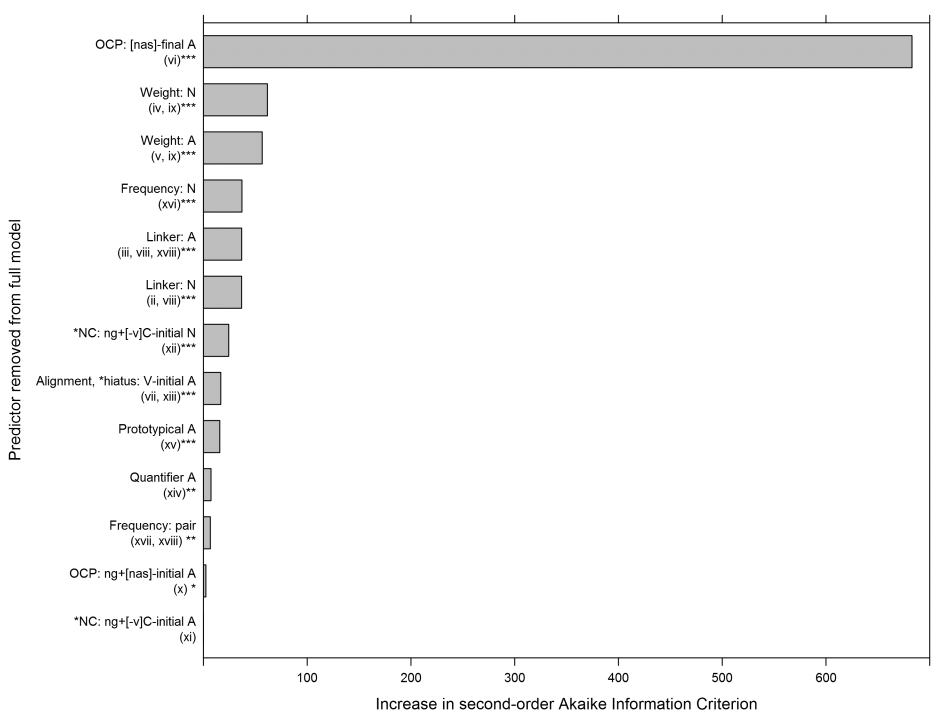


FIGURE 10. AIC_c results of predictor importance.

4.1. GENERAL PREDICTORS RESULTS.

BASIC DEFAULT-ORDER PREFERENCE. The regression model confirms the descriptive result (§3.1) that adjective + noun order is significantly more common than noun + adjective order. This is captured by the large negative intercept term (Table 2:i): when all else is held constant at the mean, there is a 91.98% probability of adjective + noun order. We find that the adjective + noun order behaves like a default order in the language. In the discussion of other factors, we will see that the adjective + noun word order is often favored: effect sizes are often larger for adjective + noun order, and adjective + noun order is often preferred when increased phonological or processing difficulty is present.

Linker preference. As noted by S&O (see §3.1), the model demonstrates a preference for the *-ng* allomorph of the linker particle over the *na* form. Holding other factors constant, the effects of linker preference are illustrated in Figure 11. The *y*-axis represents the probability of noun + adjective order (labeled here N+A), which we have plotted for the four different cases: noun + *na*, noun + *ng*, adjective + *na*, and adjective + *ng*. The error bars here and in the following plots represent 95% confidence intervals.

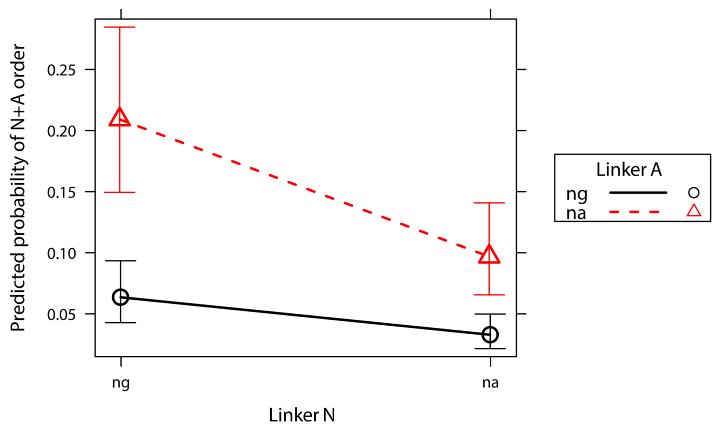


Figure 11. Effects of linker preference.

As shown in Fig. 11, there is an overall preference for the *-ng* linker. When the noun would take the *-ng* linker, the predicted probability of noun + adjective order is greater, as shown by the downward slopes of the two plotted lines. When the adjective would take the *-ng* linker, noun + adjective order is less likely, as shown by the contrast between the lower, solid line and the upper, dashed line.

We find a significant interaction effect between the linkers of the adjectives and nouns. If the noun takes *-ng* and the adjective takes *na* (e.g. *áso-ng ulól* 'dog-LINK mad' versus *ulól na áso* 'mad LINK dog'), then the probability of noun + adjective word order is 11.22% greater than if both noun and adjective took *na* linkers: this interaction is represented in Fig. 11 by the steeper slope of the dashed line (Linker A = *na*). Conversely, if the adjective takes the *-ng* linker while the noun takes the *na* linker (e.g. *bágo-ng títser* 'new-LINK teacher' versus *títser na bágo* 'teacher LINK new'), then the probability of the noun occurring first falls by 3.33%, as compared to a case in which both potential linkers are the *na* form.

The model results show that an order that promotes a *-ng* linker will be more likely to surface, all else being equal. The surface forms of both potential linkers—that of the adjective and that of the noun—compete in the model. Linker preference is shown in Fig. 10 above to be an important predictor. We find in the regression model a stronger size of effect for the adjective linker (β = 1.2793) than for the noun linker (β = −0.7744), although the importance of the adjective linker does not significantly outrank the importance of the noun linker in predictive adequacy ($\Delta AIC_c = 0.07$).

Our results of linker preference suggest that knowledge of the surface, phonologically conditioned allomorph of the linker must be present at the time of word-order choice. That we find an interaction further points to some amount of competition between the linkers of nouns and of adjectives: the linkers that both the noun and the adjective would take must be known in order to prefer *-ng*.

4.2. Phonological results. Among the phonological predictors tested, the reliable and important contributing factors include weight, adjacent-nasal avoidance, contextual markedness (*NC̥), and morphophonological alignment. Syllable-structure optimization (i.e. hiatus avoidance) was found to have a trending effect. The phonological factors that were tested and not found to influence adjective/noun order included adjacent-velar avoidance and long-distance segmental phonotactics. The reliable factors found in the model are discussed in turn below.

Weight. As with the basic descriptive statistics in §3.2, the regression model (Table 2) finds significant effects of weight, measured by segment length. What we find for the most part is that increasing the length of either word leads to more use of the default adjective + noun order.

Under the heavy-last hypothesis, we expect that longer words shift toward final position. The regression model shows this to be true for nouns: holding adjective length constant at the mean, an increase from an average-length, six-segment noun to a nine- or ten-segment noun results in a 1.2% decrease in noun + adjective order probability. That is, longer nouns have an even higher rate than normal of adjective + noun order, as expected under the heavy-last principle.

But for adjectives, we find an opposite effect. Running counter to heavy-last expectations, longer adjectives prefer adjective + noun order. Holding noun length constant, an increase from an average-length, six-segment adjective to a nine-segment adjective also decreases the probability of noun + adjective order, by 2.4%. Adjective + noun order seems to be preferred whenever the length of either the noun or adjective increases. The result that increasing adjective length also leads to adjective + noun order is surprising, given a heavy-last expectation.

A unified description of the noun and adjective effects is that having longer words results in an even stronger preference for the default order. In the case of noun length, this happens to be consistent with a heavy-last effect, but in the case of adjective length, it is not.

The model demonstrates an interaction between adjective length and noun length. When the noun is short (e.g. two to four segments long), longer adjectives increase the probability of noun + adjective order. But when the noun is longer (i.e. more than four segments), longer adjectives increase the likelihood of the default adjective + noun order.

Usual processing-based explanations of weight effects have assumed that peripheral weight effects (heavy-last or heavy-first) make it easier to balance the processing load. We hypothesize that the default-order preference exhibited by long nouns and long adjectives in Tagalog may also facilitate language processing, in that using the default order removes the cognitive uncertainty of word-order variation. When processing difficulty is compounded by the use of longer nouns or adjectives, using a default order lightens the processing load by enforcing some certainty. While such an explanation of default-order preference is consistent with uncertainty-minimization effects in psycholinguistics (e.g. Frank & Jaeger 2008), we are currently unaware of parallel phenomena to the Tagalog case.

As shown via the drop-one AIC_c test, both noun and adjective weights are robust, important predictors of variable adjective/noun word order, second only to one predictor of adjacent-nasal avoidance.

Phonotactics. Two phonotactic effects were found to be reliable and important: both were OCP effects, avoiding sequences of adjacent nasal segments (*[nasal][nasal]). Other OCP constraints that were tested—for instance, avoiding adjacent velar sequences—were not found to be reliable.

The model reveals a main effect of nasal OCP when an adjective has a final nasal, illustrated in Figure 12. When the adjective is nasal-final, adjective + noun order is much less common than usual, so as to avoid a sequence of the adjective-final nasal followed by a nasal-initial linker particle, either *-ng* or *na* (e.g. *pelúka na itím* 'wig LINK black' versus *itím na pelúka* 'black LINK wig').

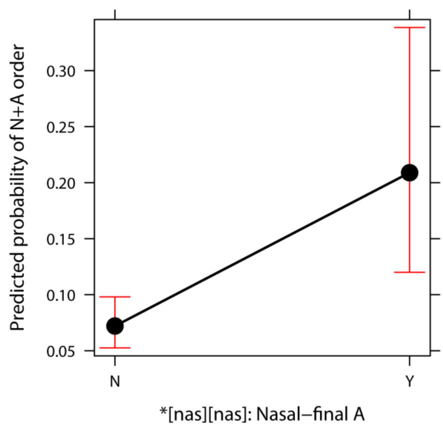


Figure 12. Effects of nasal OCP for nasal-final adjectives.

Noun + adjective order is predicted to be 13.69% more likely when an adjective is nasal-final, compared to a non-nasal-final adjective. This OCP avoidance is furthermore shown to be the most important factor in the model: the fit of a model without this factor drops drastically ($\Delta AIC_c = 682.92$). This factor is stronger than even psycholinguistic factors (e.g. frequency, weight) and semantic factors (e.g. semantic class of adjective), which are known to have strong effects on constituent ordering crosslinguistically and which normally outweigh phonological factors (see e.g. Hinrichs & Szmrecsányi 2007, Grafmiller 2014, Shih et al. 2015 on genitive construction choice in English).

This nasal OCP effect has two subcases: instances where the adjective ends in [m] or [ŋ], and instances where the adjective ends in [n]. In the latter case, there is a phonological repair for sequences of [n] followed by the linker—where the [n] is stripped, and the word takes the *-ng* linker form, but this comes at a faithfulness cost. We see here that even though the *-ng* linker form is generally preferred (§4.1), a word order that would need to strip [n] and add *-ng* is avoided. Both nasal OCP cases (with [m/ŋ]-final adjectives and [n]-final adjectives) were tested independently of each other using subset regression and/or interaction terms, and both demonstrate significantly the nasal OCP effect.[9]

[9] In a model with an interaction term between nasal-final adjective and linker form, [n]-final adjectives have a positive coefficient ($\beta = 1.166$), predicting more noun + adjective order when compared to the intercept ($\beta_0 = -2.445$). Adding the main-effect and interaction terms, adjectives with final [m] and [ŋ] have an adjusted interaction coefficient of 1.785 (i.e. coefficient $\beta = -0.674$, with the main effect coefficients for nasal-final adjectives and linker form), which is still more positive than the intercept.

The second nasal OCP effect that the model demonstrates is an avoidance of nasal-initial adjectives following the *-ng* form of the linker, shown in Figure 13.

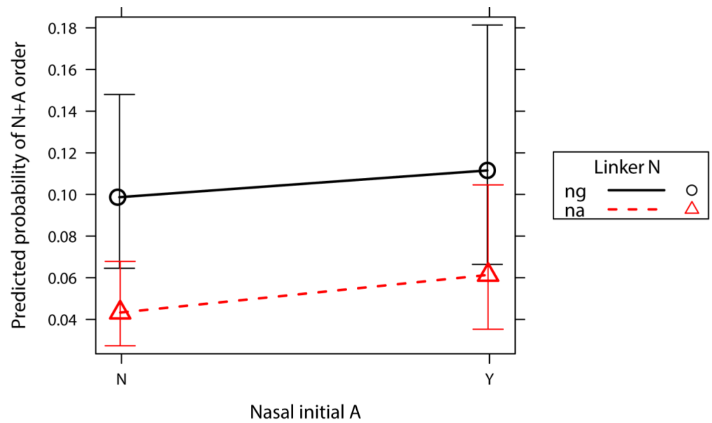


FIGURE 13. Effects of nasal OCP for nasal-initial adjectives.

As discussed above, noun + adjective order is more likely when the noun would take the *-ng* linker, as shown by the higher values of the solid line in Fig. 13 versus the dashed line. Further, noun + adjective order is more likely when the adjective begins with a nasal, seen in the upward slope of the two lines. Crucially, however, the solid line has a shallower slope: the advantage of noun + adjective order with the *-ng* linker becomes smaller when the adjective is nasal-initial, which would result in a nasal OCP violation (the noun's *-ng* would be followed by the adjective's initial nasal). As shown by the AIC test in Fig. 10, nasal-nasal avoidance in noun + adjective sequences contributes significantly less explanatory power than the same avoidance in adjective + noun sequences (i.e. factor Table 2:vi), as shown in Fig. 12. We see below that many of the effects found here are strongest when they govern well-formedness preferences in the default adjective + noun order.

CONTEXTUAL MARKEDNESS: *NC̥. Contextual markedness effects (see §3.2) were confirmed in the multivariate model results: adjective/noun orders that result in marked nasal-voiceless consonant clusters (NC̥) are avoided. In this way, word order is a viable alternative to avoiding contextually marked phonological structures in Tagalog, which in other contexts are repaired via phonological or morphological operations such as deletion and blocking.

The plot in Figure 14 demonstrates that when a noun begins with a voiceless consonant and the adjective takes the *-ng* linker (e.g. *péra-ng nakalaán* 'money-LINK dedicated' versus *NC̥-violating *nakalaá-**ng** **péra***), then noun + adjective order is more likely than when the noun begins with a voiced consonant (upward slope of the solid line). As we expect, there is no effect of consonant voicing when the adjective takes the *na* linker, as seen in the flat slope of the dashed line (e.g. neither *diyós na túnay* 'God LINK true' nor *túnay na diyós* violates *NC̥). The plot in Figure 15 demonstrates a similar pattern for cases in which the adjective begins with a voiceless consonant and the noun takes the *-ng* linker: noun + adjective order is penalized when it would produce a marked NC̥ cluster.

AIC_c comparisons in Fig. 10 indicate that avoiding NC̥ clusters that would be formed in an adjective + noun sequence is an important contributor in the model. For cases of

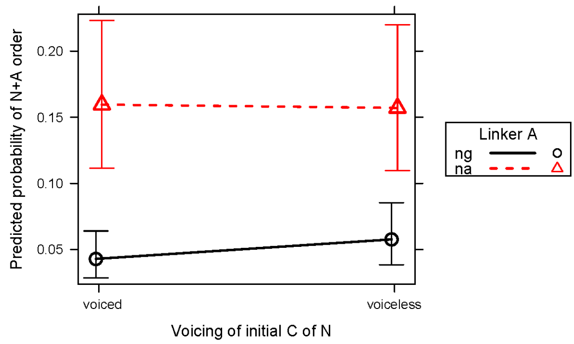


FIGURE 14. Effects of voicing of initial C of nouns.

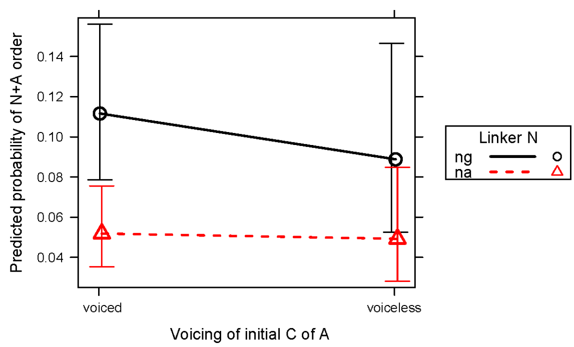


FIGURE 15. Effects of voicing of initial C of adjectives.

noun + adjective, the NC̥ avoidance effect is reliable but not as important. This AIC_c result suggests that the phonological properties of the nondefault order are not as important to word-order choice. We discuss this difference in default and nondefault ordering results further in §6.

ALIGNMENT AND SYLLABLE-STRUCTURE OPTIMIZATION. The effects of alignment and syllable-structure optimization are plotted in Figure 16. As the plot demonstrates, noun + adjective order is most likely when the noun takes the *-ng* linker and is followed by a consonant-initial adjective, which respects both syllable structure and alignment. When the noun is followed by a vowel-initial adjective, the likelihood of noun + adjective order significantly decreases: in cases where the noun takes the *-ng* linker, a following vowel-initial adjective would result in morphophonological misalignment (e.g. *pulá-**ng** itlóg* 'red-LINK egg'), and in cases where the noun takes the *na* linker, a following vowel-initial adjective would result in hiatus, requiring unfaithful [ʔ] epenthesis (e.g. *bágay **na** ˀespesyál* 'thing LINK special'). AIC_c comparisons (Fig. 10) show that the avoidance of hiatus and misalignment significantly contributes explanatory power to predicting adjective/noun order.

4.3. NONPHONOLOGICAL RESULTS.

ADJECTIVE PROTOTYPICALITY. As was expected, adjectives that begin with the prototypical *ma-* (see §3.4) are significantly biased toward adjective + noun order. Prototypicality, despite being a semantic, nonphonological predictor, is shown by the drop-one AIC_c test to contribute less explanatory importance to the model than many phonological effects.

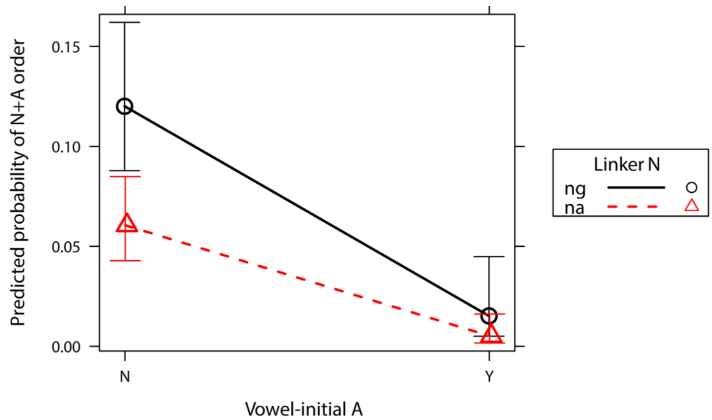


FIGURE 16. Effects of alignment and syllable-structure optimization.

QUANTIFIER ADJECTIVES. The model shows that when the adjective is a quantifier, default adjective + noun order is more likely. Taken together with prototypical adjective results, it appears that the less noun-like a given adjective is—that is, if it has the prototypical *ma-* beginning, and if it is a quantifier—the more it prefers default adjective + noun order. It is, however, notable that, like prototypicality, the quantifier status of an adjective is not as important as many phonological conditioners (see AIC_c, Fig. 10).

FREQUENCY. As noun frequency increases, the likelihood of noun-adjective order also increases. Increasing the noun's corpus frequency from the mean of approximately 1,600 to 8,000 (one standard deviation's increase on the log scale) results in a roughly 3% increase in noun + adjective order. This result accords with expectations: more-frequent words are easier to access first. From AIC results, we see that noun frequency is highly important to the explanatory power of the model, though outranked by certain phonological and processing factors including repetition avoidance (OCP) and weight.

For pair frequency, we find that higher-frequency pairs—after controlling for individual noun and individual adjective frequencies—are slightly more likely to occur in noun + adjective order, if the adjective takes the *-ng* linker. There is no effect for cases when the adjective takes *na*.

Frequency may be confounded with givenness or predictability, which we did not measure in the corpus. It would be interesting in future research to compare the contributions of frequency, predictability, and givenness.

5. COMPARISON TO A SPEECH CORPUS. We investigate here adjective/noun ordering in the only spoken corpus of Tagalog available, as a comparison to the written web corpus. We used our dictionary-derived lists of nouns and adjectives to search for adjective/noun pairs (with linker) in the Philippine Speech Corpus (Guevara et al. 2002), which includes about five minutes each of spontaneous speech from sixty-five talkers, transcribed into normal spelling.

Adjective/noun data are sparse: forty-four tokens (thirty-eight types), of which twenty-six have a quantifier adjective and eighteen a nonquantifier adjective.[10] No

[10] Looking more broadly, we did come across a few additional examples in the transcripts of noun + adjective order (but which were not found by our procedure, because either the noun or the adjective was not on the list used for the web corpus collection), such as *kapatid na nakatatandang* 'brother LINK elder'.

quantifiers occur second, and of the nonquantifiers, only one token (*báta-ng payát* 'child-LINK thin') occurs in noun + adjective order. This token is consistent with the preference of the *-ng* linker form over the *na* form: the alternative word order would produce *payát na báta* 'thin LINK child', with the less preferred *na* linker (see discussion in §4.1). Similarly, as expected by our model, sixteen of the seventeen other tokens of nonquantifier adjectives that occur in adjective + noun order respect the preference for *-ng* (e.g. observed *bágo-**ng** gámit* 'new-LINK thing' versus nonobserved *gámit **na** bágo* 'thing LINK new'). The single counterexample (observed *mataás na paaralán* 'high LINK school', versus nonobserved *paaralá-ng mataás* 'school-LINK high') exhibits a prototypical adjective, which, given our model's prediction, biases the pair toward adjective + noun ordering in spite of the *na* linker form (see e.g. §4.3).

Is the spoken data significantly different from the written data? In the written corpus, 1.9% of quantifiers take noun + adjective order, so we would expect 0.5 tokens out of twenty-six to display this order; this is not significantly different from the zero that we observe ($p = 1$ by Fisher's exact test). In the written corpus, 10.8% of nonquantifiers take noun + adjective order, and we would expect 1.9 tokens out of eighteen to display this order; this is not significantly different from the one that we observe ($p = 0.71$).

What we can conclude from this small set of data is that the spoken data are like the written data in that there is a strong overall preference for adjective + noun order, with noun + adjective order possible, too. We also observe multivariate considerations, including linker surface form and adjective prototypicality, potentially playing a role in speakers' choices of order.

6. DISCUSSION. The investigation presented here examined how phonology influences word-order variation in Tagalog. The major results are discussed in turn.

First, our corpus study reveals that, controlling for nonphonological predictors, phonological well-formedness conditions significantly contribute to predicting variable adjective/noun order, an alternation that had previously been characterized as primarily 'free'. This is particularly true for the overwhelmingly preferred adjective + noun order, even once controls for individual words and frequencies are included. This quantitative result aligns with observations in the previous Tagalog literature that one order is more basic—or, DEFAULT—than the other (see discussion in §§3.1 and 4.1). The conditioning factors with the strongest effect sizes tend to be those that regulate phonological well-formedness of the adjective + noun order (versus the noun + adjective order): that is, phonological ill-formedness in the DEFAULT order appears to be more strictly penalized, resulting in the use of the alternative order. For example, although the surface form of the linker for both adjectives and nouns contributes to order choice, the linker form that the adjective would take, if it comes first, has a stronger effect (adjective linker: $\beta = 1.2793$, versus noun linker: $\beta = -0.7744$). Similarly, nasal-nasal avoidance (OCP; see results in §4.2) is enforced more strongly when an adjective is nasal-final (so that adjective + noun order produces a nasal-nasal sequence: $\beta = 1.224$) than when an adjective is nasal-initial and the noun takes the *-ng* linker (so that noun + adjective order produces a nasal-nasal sequence: $\beta = 0.224$). The results also show a stronger effect of *N$\mathring{C}$ in the default adjective + noun order, where a voiceless consonant-initial noun would form an illicit N$\mathring{C}$ cluster with an *-ng* linker on the adjective ($\beta = -0.3299$, versus $\beta = 0.1992$ for voiceless-C-initial adjective and *-ng* linker on the noun; see §4.2). One possible exception to this otherwise strong pattern is alignment and hiatus avoidance, where our model shows an effect for vowel-initial adjectives (penalizing noun + adjective order) but not vowel-initial nouns (penalizing adjective + noun order). The finding that the default order is more phonologically regulated demonstrates that phonological considerations and morphosyntactic preferences interact: nonoptimal phonological structure can, in some cases, outweigh otherwise strong default preferences of word order.

Second, we find that the SURFACE form of the linker morpheme plays a role. The *-ng* form of the linker is preferred to the *na* form (§4.1). Additionally, the surface form of the linker is used in assessing the phonological conditions of nasal OCP, contextual markedness, and hiatus avoidance (§4.2)—all of which are shown to significantly predict adjective/noun order. This result suggests that the surface phonological form must be available for word-order choice.

Finally, our corpus investigation illuminates the phonological factors that are most likely to interact with morphosyntactic choices. To date, one major reason many language models limit the influence that phonology can have in morphosyntactic phenomena is the apparent empirical rarity of such phonology-morphosyntax behaviors (e.g. Zwicky & Pullum 1986a,b). Previous studies of constituent order have demonstrated some effects of phonological influence in syntactic variation (e.g. McDonald et al. 1993, Wright et al. 2005, Benor & Levy 2006, Mollin 2012, Shih et al. 2015), but the existing literature has largely focused on a limited set of phonological factors repeatedly shown to influence constituent-order variation: that is, prosody and weight (e.g. Zec & Inkelas 1990, Zubizarreta 1998, Anttila et al. 2010, Ehret et al. 2014). Individual studies have also demonstrated a smattering of segmental effects: adjacent segmental-identity avoidance (e.g. adjacent-sibilant avoidance in English genitive constructions: Hinrichs & Szmrecsányi 2007); long-distance phonotactic preferences (e.g. high to low vowel-height ordering in English:[11] *teeny tiny* versus ???*tiny teeny*: Cooper & Ross 1975; though cf. Benor & Levy 2006); and syllable-structure optimization (e.g. hiatus avoidance in English binomial order: Wright et al. 2005). For surveys of such phenomena, see, for example, Wolf 2008, Shih 2014, 2017, Anttila 2016.

The crucial question that has yet to be answered is WHICH phonological conditions affect constituent ordering. Many phonological factors that have been investigated in word-order studies arose from incidental observation. But are there uniting principles behind what the relevant phonological factors are? Understanding the uniting principles, if any, behind which phonological factors can affect word order is key to grasping how much phonology interacts with syntactic variation, and, for future work, key to systematically identifying the crucial phonological factors from the grab bag of possibilities. Perhaps phonological influences on morphosyntactic variation are not as rare as previously thought if we know where to look.

It has previously been proposed that phonological influences on constituent-order variation should be limited to prosodic conditions, including weight (prosodically defined), phrasal stress, and intonation (e.g. Selkirk 1984, Inkelas 1990, Zec & Inkelas 1990). The idea is that higher-level prosodic components interface (in a limited fashion) with syntactic information. Lower-level segmental phonology, by contrast, interfaces solely with prosody, not syntax, thereby maintaining some unidirectionality in the linguistic system (see also e.g. Zwicky & Pullum 1986a,b, Levelt 1989, Vogel & Kenesei 1990, Bock & Levelt 1994, Miller et al. 1997, Garrett 2000, Scheer 2012). Hence, under such a view, only prosodic information has the ability to participate in the ordering of syntactic constituents, and segmental effects (e.g. OCP, contextual markedness) are a priori excluded.

[11] For a non-English example, see for example Mortensen 2006:222–23 for vowel-height preferences in Jingpho compound coordination.

Given that segmental effects that influence constituent ordering have been identified in the current study and elsewhere, the view that we present here departs from the narrower assumption that phonological conditions on word order are solely limited to the domain of prosody (see also Wolf 2008:240 for similar claims regarding the influence of phonology on morpheme order). Our working proposal instead is that the phonological conditioning factors that play a role in influencing word-order variation are the ones that also affect syntagmatic configurations elsewhere in the phonological grammar of the language (see also e.g. Shih 2014). On the assumption that the same phonological conditions that govern phonotactics and (morpho)phonological alternations (within words) will be active in the grammar of a language at large (e.g. Kenstowicz & Kisseberth 1977, Raffelsiefen 1999, Martin 2011), we expect then that the conditions that receive support from lexical and morphophonological patterns will exhibit stronger effects on word-order choice than those that do not. Evidence that the speaker and language learner get from within-word phonological patterns will tend to generalize into across-word domains in the (probabilistic) grammar of a language (Martin 2011). In this way, phonology-syntax interaction is constrained not by top-down, theoretically imposed limitations but instead by evidence to the learner.

Our view, then, allows word-order conditioning factors to come from the prosodic domain AND from the rest of the phonological system. We predict that the same well-formedness constraints on phonological strings that trigger repairs or avoidances within words in a language are likely to be the ones that trigger repairs and avoidances across words, as elements (i.e. words, constituents, phrases) combine. The repair strategy used in the case explored here is word order: when an illicit phonological structure is encountered at the combination of a noun and an adjective, a word order that avoids the ill-formedness is more likely to surface than it would have been had the alternative order not resulted in phonological ill-formedness. Syntagmatic phonological conditions such as contextual markedness and the OCP, which are familiarly satisfied via phonological optimization, can also be satisfied via word order.

The two local OCP conditions that we tested in this study—*[nasal][nasal] and *[velar][velar]—illustrate our prediction. As discussed in §3.2, adjacent clusters of [nasal] segments are underrepresented in the lexicon. In comparison, the picture is different for velar-velar sequences. In 497 consonant clusters in the lexicon, the first consonant is velar, and in 165 clusters, the second consonant is velar. We therefore expect sixty-five velar-velar clusters if consonants combined freely, and in fact, there are 152—many more than expected. We find the parallel result in our study of adjective/noun word order: orderings that result in adjacent nasals are avoided (e.g. *[nasal]-final adjective + linker), but there is no reliable effect of velar-cluster avoidance (e.g. *-*ng*-linked adjective + velar-initial noun: *kawáwa-**ng gúro*** ‘poor teacher’). In other words, a phonotactic constraint that is evident within words in Tagalog also appears to govern phonotactic preferences across word boundaries, and a phonotactic constraint that has little evidence from lexical patterns is not strongly present across word boundaries.[12] Note that avoiding identical-place segments does happen crosslinguisti-

[12] A referee points out that the OCP as normally understood may not suffice to explain the observed pattern. The difference between [velar][velar] and [nasal][nasal] sequences within a word would typically be explained by saying that successive velars can share their place feature—thus escaping an OCP violation—and successive nasals cannot, for some reason, share their [nasal] feature. Feature sharing would not be expected across a word boundary, so a sequence of velars across a word boundary SHOULD be penalized. Or more precisely, the abundance of [velar][velar] sequences within words should not imply that these sequences are also licit across words. We see two options to explain the observed pattern. First, it could be that feature sharing IS

cally. For example, Arabic verb roots are poor in sequences of dorsals as compared to sequences of nasals (e.g. OCP-PLACE >> OCP-MANNER) (e.g. McCarthy 1979, Frisch et al. 2004). So, the parallelisms we find between within-word and between-word patterns in Tagalog do not arise from the crosslinguistic absence of [velar][velar] avoidance.

The phonological effects that we found to be important and reliable in predicting adjective/noun order are ones that are active in the phonological grammar, in either phonotactics or alternations. For example, within words, the penalty on a nasal followed by a voiceless consonant (*NC̥) causes deletion or fusion. Another example is morpheme-to-syllable alignment. We saw its effect at the adjective-noun boundary, and it also applies within words, where syllabification respects the prefix-stem boundary, as in [mag-] + [ˀalis] 'departure' → [magˀalis] 'remove', *[magalis] (example from S&O p. 19), even though within a morpheme, consonant + [ʔ] sequences are extremely rare.

We summarize in Table 6 the phonological factors tested in our regression model (§4; omitting weight, which is not relevant within words) and compare their activity in adjective/noun ordering to their activity within single words.[13]

FACTOR	ACTIVE WITHIN WORDS?	ACTIVE IN ADJECTIVE/NOUN ORDER?
Nasal OCP	YES • [nasal][nasal] underrepresented in roots	YES • [...[nasal]$_{\text{adjective}}$] *na* noun penalized • noun *na* [$_{\text{adjective}}$[+nasal]...] penalized
Velar OCP	NO • [velar][velar] not underrepresented in roots	NO • no significant effect (eliminated from regression model)
*NC̥	YES • [nasal][−voice] underrepresented in roots • Prefix-final [nasal] fuses with stem-initial [−voice].	YES • noun-*ng* [$_{\text{adjective}}$[−voice]...] penalized • adjective-*ng* [$_{\text{noun}}$[−voice]...] penalized
Alignment: *C + (ʔ)V	UNCLEAR • No active alternation, but allows a sequence that is phonotactically illegal elsewhere within a word. One could investigate whether consonant-final prefixes attach less often than expected to vowel-initial stems.	YES • noun-*ng* [$_{\text{adjective}}$ V...] penalized
Syllable structure: *V(ʔ)V	YES • Vowels within a root are separated by [ʔ], though in casual speech this is optionally deleted (S&O:19).	YES • noun *na* [$_{\text{adjective}}$ V...] penalized

TABLE 6. Phonological factors within words versus in adjective/noun ordering.

On our approach, allowing segmental phonology to influence word order does not mean that ANY segmental information can willy-nilly influence word order. Zwicky and Pullum (1986a:75) offer the classically cited example of a phonology-syntax interaction that is not expected to exist (their emphasis): 'a movement transformation that obligatorily moves … [a] constituent that begins *phonetically* with a bilabial consonant', maintaining that surface segmental information should not affect word order. However, the explanation we pursue from the Tagalog evidence observed here is not that segment-

possible across a word boundary (and for the same reasons as word-internally, [velar] can be shared but [nasal] cannot), even for the underlyingly velar linker and a following underlyingly velar consonant. Or second, we can assume that rather than the OCP (a prohibition on successive, distinct autosegments), we are dealing with simpler constraints: one against successive nasal consonants, regardless of their structure, and one against successive velars, regardless of structure.

[13] '+' denotes a morpheme boundary.

induced word-order differences are a priori ruled out by the grammar, but rather that segment-induced word-order variation is unlikely without motivation. For example, suppose that there were a condition, active in the phonology of a language, against two labials in succession (an instance of the OCP; Leben 1973): we would predict a greater likelihood of syntactic repairs such as fronting of, for example, [*si**b***] [***p**al*] to [*pal*] [*sib*] (see also Shih 2014:213ff.).[14] This is what the Tagalog data shows: the same phonological principles that drive morphophonological alternations and phonotactic patterns are the ones that also affect word-order variation. We maintain here that only the phonological constraints present crosslinguistically and most especially those with supporting evidence from the phonotactic and (morpho)phonological alternation patterns in a given language will be the ones that can condition word-order alternations as well.

7. Conclusion. In this article, we presented a case study of Tagalog adjective/noun word order, demonstrating that word-order variation, which was previously characterized as 'free', can be predicted from a number of syntactic, semantic, and phonological properties. We focused in particular on identifying the phonological predictors relevant to word-order choice. Our results suggested that word order is conditioned by both prosodic and segmental phonological conditions that are extant elsewhere within the grammar of Tagalog, producing surface word orders that are phonologically optimal. In short, our evidence showed that certain syntactic behaviors, such as word-order variation, are not entirely free from the influence of phonology.

Appendix A: Excluded items due to ambiguity between target item and common nontarget item

EXCLUDED ITEM	REASON
alam	noun 'knowledge' or verb 'know'
am	noun 'rice broth' or English *A.M.*
dapat	adjective 'worthy' or verb 'should'
habang	noun 'length' or preposition 'while'
hanggang	noun 'result' plus linker or preposition 'until'
lamang	noun 'advantage' or particle 'only'
pain	noun 'bait' or English *pain*
said	adjective 'consumed' or English *said*
sayang	noun 'long skirt' or particle 'what a pity'
silang	noun 'mountain pass' or pronoun 'they' plus linker
tapos	noun 'ending' or preposition 'after'
todo(ng)	listed in dictionary as *adj. 'all'*, but seems to function as quantifier
upang	noun 'rent' plus linker or conjunction 'in order to'

Appendix B: Quantifier adjectives

LIMITERS (string in corpus)

bahagi	'part (of)'	hindi	negative particle
bawat	'each, every'	iilan	'only a few'
kaunti	'a little, little, few'	ilan	'a few, some, certain'
kakaunti	'only a little'	huli	'(the) last'
kalahati	'half (of)'	lahat	'all'
kapiraso	'a piece of'	marami	'a lot (of), many'
kaputol	'a piece of'	parte	'part (of)'
karamihan	'most (of), majority (of)'		

[14] Moving a constituent with a bilabial consonant, as in Zwicky and Pullum's example, would not repair a simple restriction against labial consonants, but if the restriction is against labial consonants in some sort of syntagmatic configuration, such as *[labial][labial], then changing the word order can fix the problem, by removing the illicit sequence of [labial] consonants. In this way, word-order alternatives from the syntax can be candidates for phonological optimization just as phonological repairs are.

NUMERALS (string in corpus)

isa	'one'
dalawa	'two'
tatlo	'three'
apat	'four'
lima	'five'
anim	'six'
pito	'seven'
walo	'eight'
siyam	'nine'
pu	'times ten'
labing[15]	'ten plus'
libo	'times one thousand'
milyon	'times one million'
sampu	'ten'
dalawampu	'twenty'
tatlumpu	'thirty'
apatnapu	'forty'
limampu	'fifty'
animnapu	'sixty'
pitumpu	'seventy'
walumpu	'eighty'
siyamnapu	'ninety'
una	'first'
ikalawa	'second'
ikatlo	'third'
ikaapat	'fourth'
ikalima	'fifth'
ikaanim	'sixth'
ikapito	'seventh'
ikawalo	'eighth'
ikasiyam	'ninth'
ikasampu	'tenth'
ikalabingisa	'eleventh'
ikadalawampu	'twentieth'
ikaisangdaan	'one hundredth'
pagunahin	'first, foremost'
paguna	'first'
pangalawa	'second'
pangatlo	'third'
pangapat	'fourth'
panlima	'fifth'
panganim	'sixth'
pampito	'seventh'
pangwalo	'eighth'
pansiyam	'ninth'
pansampu	'tenth'
panlabingisa	'eleventh'
pandalawampu	'twentieth'
pangisangdaan	'one hundredth'

REFERENCES

Anttila, Arto. 2016. Phonological effects on syntactic variation. *Annual Review of Linguistics* 2.115–37. DOI: 10.1146/annurev-linguistics-011415-040845.

Anttila, Arto; Matthew Adams; and Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25.946–81. DOI: 10.1080/01690960903525481.

Baroni, Marco; Silvia Bernardini; Adriano Ferraresi; and Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43.209–26. DOI: 10.1007/s10579-009-9081-4.

Bartoń, Kamil. 2013. MuMIn: Multi-model inference. R package. Online: http://cran.r-project.org/web/packages/MuMIn/index.html.

Bates, Douglas; Martin Maechler; and Ben Bolker. 2013. lme4: Linear mixed-effects models using 'Eigen' and S4. R package. Online: https://cran.r-project.org/web/packages/lme4/.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25.110–42. DOI: 10.1515/9783110242652.110.

Bennett, William G. 2013. *Dissimilation, consonant harmony, and surface correspondence*. New Brunswick, NJ: Rutgers University dissertation.

Benor, Sarah Bunin, and Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82.233–78. DOI: 10.1353/lan.2006.0077.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas; Stig Johansson; Susan Conrad; and Edward Finnegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Bock, Kathryn, and Willem Levelt. 1994. Language production: Grammatical encoding. *Handbook of psycholinguistics*, ed. by Morton A. Gernsbacher, 945–84. New York: Academic Press.

Burnham, Kenneth P., and David R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd edn. New York: Springer.

[15] *labi*, without the linker *ng* (as listed in the SEAsite dictionary), is also a noun, meaning 'lips'. The limiter definition in SEAsite for *labing* is 'excess'.

Burnham, Kenneth P., and David R. Anderson. 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* 33.261–304. DOI: 10.1177/0049124104268644.

Cohen Priva, Uriel. 2010. Constructing typing-time corpora: A new way to answer old questions. *Proceedings of the 32nd annual meeting of the Cognitive Science Society (CogSci 2010)*, 43–48.

Cooper, William E., and John R. Ross. 1975. World order. *Chicago Linguistic Society (Parasession on functionalism)* 11(2).63–111.

Donohue, Mark. 2007. Word order in Austronesian from north to south and west to east. *Linguistic Typology* 11.349–91. DOI: 10.1515/LINGTY.2007.026.

Ehret, Katherina Luisa; Christoph Wolk; and Benedikt Szmrecsányi. 2014. Quirky quadratures: On rhythm and weight as constraints on genitive variation in an unconventional data set. *English Language & Linguistics* 18.263–303. DOI: 10.1017/S1360674314000033.

English, Leo. 1986. *Tagalog-English dictionary*. Manila: Congregation of the Most Holy Redeemer; distributed by Philippine National Book Store.

Fox, John. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8.1–27. DOI: 10.18637/jss.v008.i15.

Frank, Austin, and T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th annual meeting of the Cognitive Science Society (CogSci 2008)*, 939–44. Online: http://escholarship.org/uc/item/7d08h6j4#page-1.

French, Koleen Matsuda. 1988. *Insights into Tagalog: Reduplication, infixation, and stress from nonlinear phonology*. (Publications in linguistics 84.) Dallas: Summer Institute of Linguistics and The University of Texas at Arlington.

French, Koleen Matsuda. 1991. Secondary stress in Tagalog. *Oceanic Linguistics* 30.157–78. DOI: 10.2307/3623086.

Frisch, Stefan A.; Janet B. Pierrehumbert; and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228. DOI: 10.1023/B:NALA.0000005557.78535.3c.

Garrett, Merrill F. 2000. Remarks on the architecture of language processing systems. *Language and the brain: Representation and processing* (Foundations in neuropsychology), ed. by Yosef Grodzinsky, Lewis P. Shapiro, and David Swinney, 3–69. San Diego: Academic Press.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27.2865–73. DOI: 10.1002/sim.3107.

Gentner, Donald R.; Serge Larochelle; and Jonathan Grudin. 1988. Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology* 20.524–48. DOI: 10.1016/0010-0285(88)90015-1.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, ed. by Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, 95–126. Cambridge, MA: MIT Press.

Goldsmith, John. 1976. *Autosegmental phonology*. Cambridge, MA: MIT dissertation.

Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language & Linguistics* 18.471–96. DOI: 10.1017/S1360674314000136.

Griffin, Zenzi M., and Kathryn Bock. 1998. Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language* 38.313–38. DOI: 10.1006/jmla.1997.2547.

Guevara, Rowena Cristina L.; Melvin Co; Evan Espina; Ian Dexter Garcia; Emerson Tan; Ryan Ensomo; and Ramil Sagum. 2002. Development of a Filipino speech corpus. Philippines, ms.

Hansson, Gunnar Ólafur. 2001. *Theoretical and typological issues in consonant harmony*. Berkeley: University of California, Berkeley dissertation.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce, and Tanya Stivers. 1996. The phonetics of post-nasal voicing. Los Angeles: University of California, Los Angeles, ms.

Hinrichs, Lars, and Benedikt Szmrecsányi. 2007. Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language & Linguistics* 11.437–74. DOI: 10.1017/S1360674307002341.

Inkelas, Sharon. 1990. *Prosodic constituency in the lexicon*. (Outstanding dissertations in linguistics.) New York: Garland.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Kelly, Michael H.; J. Kathryn Bock; and Frank C. Keil. 1986. Prototypicality in a linguistic context: Effects on sentence structure. *Journal of Memory and Language* 25.59–74. DOI: 10.1016/0749-596X(86)90021-5.

Kenstowicz, Michael, and Charles Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Leben, William R. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT dissertation.

Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Maclay, Howard, and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15.19–44. DOI: 10.1080/00437956.1959.11659682.

Martin, Andrew T. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87.751–70. DOI: 10.1353/lan.2011.0096.

McCarthy, John J. 1979. *Formal problems in Semitic phonology and morphology*. Cambridge, MA: MIT dissertation.

McCarthy, John J., and Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology* 1993.79–153. DOI: 10.1007/978-94-017-3712-8_4.

McDonald, Janet L.; Kathryn Bock; and Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25.188–230. DOI: 10.1006/cogp.1993.1005.

Menn, Lise, and Brian MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language* 60.519–41. DOI: 10.2307/413990.

Miller, Philip H.; Geoffrey K. Pullum; and Arnold M. Zwicky. 1997. The principle of phonology-free syntax: Four apparent counterexamples in French. *Journal of Linguistics* 33.67–90.

Mollin, Sandra. 2012. Revisiting binomial order in English: Ordering constraints and reversibility. *English Language & Linguistics* 16.81–103. DOI: 10.1017/S1360674311000293.

Mortensen, David. 2006. *Logical and substantive scales in phonology*. Berkeley: University of California, Berkeley dissertation.

Nolasco, Ricardo. 2007. Tagalog and Filipino: Not so simple/How to value our languages. *dalityapi unpoemed*. Online: http://svillafania.philippinepen.ph/2007/08/articles-filipino-and-tagalog-not-so.html.

Nunberg, Geoffrey. 1990. *The linguistics of punctuation*. (CSLI lecture notes 18.) Stanford, CA: CSLI Publications.

Onishi, Kristine H.; Gregory L. Murphy; and Kathryn Bock. 2008. Prototypicality in sentence production. *Cognitive Psychology* 56.103–41. DOI: 10.1016/j.cogpsych.2007.04.001.

Pater, Joe. 1996. *NC. *North East Linguistic Society (NELS)* 26.227–39.

Pater, Joe. 2001. Austronesian nasal substitution revisited: What's wrong with *NC (and what's not). *Segmental phonology in optimality theory: Constraints and representations*, ed. by Linda Lombardi, 159–82. Cambridge: Cambridge University Press.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. RuCCS Technical Report. New Brunswick, NJ: Rutgers University, ms.

Quirk, Randolph; Sidney Greenbaum; Geoffrey Leech; and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. *Yearbook of Morphology* 1998.225–87. DOI: 10.1007/978-94-017-3720-3_10.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement by correspondence. *Language* 80.475–531. DOI: 10.1353/lan.2004.0144.

Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81.613–44. DOI: 10.1353/lan.2005.0149.

Ryan, Kevin M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28.413–54. DOI: 10.1017/S0952675711000212.
Ryan, Kevin M. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90.309–41. DOI: 10.1353/lan.2014.0029.
Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R*. New York: Springer.
Schachter, Paul, and Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
Scheer, Tobias. 2012. *Direct interface and one-channel translation: A non-diacritic theory of the morphosyntax-phonology interface* (*A lateral theory of phonology, vol. 2*). Berlin: De Gruyter Mouton.
SEAsite. 2001. SEAsite Tagalog online dictionary (Large Tagalog dictionary version). Online: http://www.seasite.niu.edu/Tagalog/on_site_dictionary.htm.
Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Shih, Stephanie S. 2014. *Towards optimal rhythm*. Stanford, CA: Stanford University dissertation.
Shih, Stephanie S. 2017. Phonological influences in syntactic choice. *The morphosyntax-phonology connection: Locality and directionality at the interface*, ed. by Vera Gribanova and Stephanie S Shih, 223–52. Oxford: Oxford University Press.
Shih, Stephanie S., and Jason Grafmiller. 2011. Weighing in on end weight. Paper presented at the 85th annual meeting of the Linguistic Society of America, Pittsburgh, PA.
Shih, Stephanie S.; Jason Grafmiller; Richard Futrell; and Joan Bresnan. 2015. Rhythm's role in predicting genitive alternation choice in spoken English. *Rhythm in cognition and grammar: A Germanic perspective*, ed. by Ralf Vogel and Ruben van de Vijver, 207–34. Berlin: De Gruyter Mouton.
Szmrecsányi, Benedikt. 2004. On operationalizing syntactic complexity. *Les Journées internationales d'Analyse statistique des Données Textuelles* 7.1031–38.
Szmrecsányi, Benedikt, and Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. *The dynamics of linguistic variation: Corpus evidence on English past and present*, ed. by Terttu Nevalainen, Irma Taavitsinen, Päivi Pahta, and Minna Korhonen, 291–310. Amsterdam: John Benjamins.
Teeple, David. 2008. Prosody can outrank syntax. *West Coast Conference on Formal Linguistics (WCCFL)* 26.454–62. Online: http://www.lingref.com/cpp/wccfl/26/paper1702.pdf.
Temperley, David. 2006. Minimization of dependency length in written English. *Cognition* 105.300–333. DOI: 10.1016/j.cognition.2006.09.011.
Topintzi, Nina. 2006. *Moraic onsets*. London: University College London dissertation.
Tranel, Bernard, and Francesca Del Gobbo. 2002. Local conjunction in Italian and French phonology. *Romance phonology and variation: Selected papers from the 30th Linguistic Symposium on Romance Languages*, ed. by Caroline R. Wiltshire and Joaquim Camps, 191–218. Amsterdam: John Benjamins.
Tucker, Matthew A., and Robert Henderson. 2010. Latest insertion in K'ichee'. Paper presented at the 84th annual meeting of the Linguistic Society of America, Baltimore, MD.
Vinson, David W.; Rick Dale; Stephanie S. Shih; and Michael J. Spivey. 2017. Lexical effects on keystroke-level dynamics reflect word predictability. Merced: University of California, Merced, ms.
Vogel, Irene, and István Kenesei. 1990. Syntax and semantics in phonology. *The phonology-syntax connection*, ed. by Sharon Inkelas and Draga Zec, 339–64. Chicago: University of Chicago Press.
Wasow, Tom. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.
Weingarten, Rüdiger; Guido Nottbusch; and Udo Will. 2004. Morphemes, syllables, and graphemes in written word production. *Language production*, ed. by Thomas Pechmann and Christopher Habel, 529–72. Berlin: Mouton de Gruyter.
Wolf, Matthew. 2008. *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. Amherst: University of Massachusetts, Amherst dissertation.

Wright, Saundra K.; Jennifer Hay; and Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43.531–61. DOI: 10.1515/ling.2005.43.3.531.

Yamashita, Hiroko, and Franklin Chang. 2001. 'Long before short' preference in the production of a head-final language. *Cognition* 81.B45–B55. DOI: 10.1016/S0010-0277(01)00121-4.

Zec, Draga, and Sharon Inkelas. 1990. Prosodically constrained syntax. *The phonology-syntax connection*, ed. by Sharon Inkelas and Draga Zec, 365–78. Chicago: University of Chicago Press.

Zesiger, Pascal; Jean-Pierre Orliaguet; Louis-Jean Boë; and Pierre Mounoud. 1994. The influence of syllabic structure in handwriting and typing production. *Advances in handwriting and drawing: A multidisciplinary approach*, ed. by Claudie Faure, P. J. G. Keuss, G. Lorette, and Annie Vinter, 389–401. Paris: Europia.

Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. (Linguistic Inquiry monographs 33.) Cambridge, MA: MIT Press.

Zuraw, Kie. 2006. Using the web as a phonological corpus: A case study from Tagalog. *EACL-2006: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics/Proceedings of the 2nd International Workshop on Web As Corpus*, 59–66.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28.417–72. DOI: 10.1007/s11049-010-9095-z.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23.133–48. DOI: 10.1017/S0022226700011063.

Zwicky, Arnold M., and Geoffrey Pullum. 1986a. The principle of phonology-free syntax: Introductory remarks. *Ohio State University Working Papers in Linguistics (Interfaces)* 32.63–91.

Zwicky, Arnold M., and Geoffrey Pullum. 1986b. Two spurious counterexamples to the principle of phonology-free syntax. *Ohio State University Working Papers in Linguistics (Interfaces)* 32.92–99.

Shih
USC Department of Linguistics
3601 Watt Way
Grace Ford Salvatori 301
Los Angeles, CA 90089-1693
[shihs@usc.edu]

Zuraw
UCLA Department of Linguistics
3125 Campbell Hall, Box 951543
Los Angeles, CA 90095
[kie@ucla.edu]

[Received 25 March 2015;
revision invited 20 December 2015;
revision received 14 August 2016;
revision invited 1 January 2017;
revision received 13 February 2017;
accepted 21 May 2017]