Phonological conditions on variable adjective-noun word order in Tagalog

Stephanie S Shih
University of California, Merced

Kie Zuraw
University of California, Los Angeles

«DRAFT under review, submitted March 2015»

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 paper presents a large-scale, web-based 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

1 INTRODUCTION

Tagalog adjective-noun word order is traditionally described as freely variable. Both orders—prenominal adjective (1a) or postnominal adjective (1b)—are claimed to be possible and semantically interchangeable, "without any apparent difference in meaning" (Schachter and Otanes 1972:121):

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

          b.   babáe-ng   magandá
                woman-LINK   beautiful
                'beautiful woman'

Despite the general claims of free order and semantic interchangeability, some brief hypotheses gleaned from casual observation regarding the conditioning of order variation have been put forth in the previous literature. Schachter & Otanes, for example, point to semantic factors such as givenness or adjective semantic class that tend to prefer one order over another. Donohue (2007:360) instead suggests that phonological factors—rather than syntactic ones—condition the word order choice.
       Tagalog adjective-noun word order variation presents an additional wrinkle in that a linker particle intervenes between the two elements (noun, adjective) to mark their relationship in the

noun phrase. 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 spelled as a separate word. The form *–ng* occurs following vowel-final words (2a). In words that have a word-final alveolar nasal [n] or glottal stop [ʔ], the [n] or [ʔ] is stripped, and the linker *–ng* attaches (2b). All other consonant-final words combine with the form, *na* (2c).

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

Schachter & Otanes (122) suggest that word order variation also interacts with the phonologically-conditioned linker in that, for a given adjective and noun pair, an order in which the *–ng* linker would surface is preferred. Such a preference would make word order variation dependent on the allomorphic *surface* phonological form

    Cross-linguistic studies of word order and construction variation have yielded evidence that morphosyntactic variation of this type is conditioned on the basis of numerous factors, including psycholinguistic, sociolinguistic, semantic, and morphosyntactic structure-based pressures. In comparison, phonological factors are generally considered to be not quite as important in conditioning word order variation. Some point to higher-level prosodic conditioning of syntactic choices, but lower-level segmental, phonotactic conditioning is left largely untouched, on the broad assumption that such minute detail of phonological form cannot affect syntactic structures.

    Given the observations from the previous literature (e.g., S&O; Donohue), Tagalog noun-adjective word order variation appears to be a case in which phonological form affects morphosyntactic variation. In this paper, we present the first systematic, large-scale empirical study of adjective and noun word order variation in Tagalog. We focus in particular on phonological factors that may affect order. Using a corpus of written Tagalog web-text (Zuraw 2006), we provide evidence that, in addition to semantic and usage-based factors, phonological well-formedness conditions influence adjective-noun word order in Tagalog. Like S&O, we find that word order is conditioned in part by the phonological surface form of the linker particle. The results presented herein demonstrate that word order can be leveraged to optimize phonological output well-formedness preferences, and that, in order to do so, surface phonological information has to be available at the point of these morphosyntactic word order choices.

    The paper is organized as follows. Details of the data are introduced in §2. In §3, we present the phonological and non-phonological conditioning factors that are included in the study. The factors investigated were drawn from both the Tagalog literature and cross-linguistic studies of morphosyntactic variation. Results of the web-based corpus study are presented in §4, and §5 includes a brief comparison of results to patterns in a small speech corpus of Tagalog. At the end (§0), we discuss the ramifications of the results, focusing on which phonological factors are most likely to affect morphosyntactic variation, and what such interaction between phonological optimization and morphosyntactic order reveals about a linguistic model of the morphosyntax-phonology interface.

2 DATA

We used a corpus of Tagalog/Filipino web text (Zuraw 2006). 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 fairly well, with each letter corresponding to one sound in most cases (though loans and proper names can have idiosyncratic spelling). Letter values are close to 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 paper we use acute accent marks to mark stress.

Fortunately, it is easy to see from a word's spelling which linker allomorph it should take, what sound it begins or ends with (except [ʔ]), how many syllables it has (one per vowel letter), and how many segments long it is.

To obtain a sub-corpus 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 and coding. 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; there was punctuation anywhere within the sequence; the item contained any of the following words, ambiguous between being a target noun or adjective and a common non-target item:

(3)   *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'

We selected 1,205 noun/adjective pairs (types) that included the nouns and adjectives that occurred most frequently in the data, as well as the most-frequent pairs, and three Tagalog-English bilinguals hand-checked these items. This checking identified 11 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*'. Adjective/noun pairs containing any of these problematic words were excluded. We were left with 149,689 adjective/noun pair tokens, representing 14,591 types. The pairs included 1,708 different nouns and 587 different adjectives. For each of the 14,591 items, we compiled the number of tokens with adjective-linker-noun order, and the number with noun-linker-adjective order, and also derived from that the item's rate of taking noun-linker-adjective order.

There are three potential sources of noise we are aware of but 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 (4), where the adjective *armado* could be part of a much longer modifier.

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

      b.   **táo-ng armado** *sa kanilang ika-apat na henersyon mobile na aprato*
           'person-LINK armed with their fourth LINK generation mobile LINK device'

Second, we don't know which of our tokens are being used as predicates (e.g., predicate *This is a **beautiful place*** versus non-predicate *Then we arrived at a **beautiful place***). Schachter & Otanes (1972:121–122) suggest that "unmarked" predicates exhibit special behavior with respect to ordering preferences. For Schachter & Otanes, "unmarked" refers to when the adjective and noun pair acts as a predicate and does not include a personal pronoun, deictic pronoun, or personal name. An example is *doktor na bantog* 'famous LINK doctor' (Schachter and Otanes 1972:121):

(5)   Doktor na   bantong ang ama   ni Juan
      doctor LINK famous ANG father NG Juan
      'Juan's father is a famous doctor.'

Schachter & Otanes state that noun-first order is preferred for unmarked predicates when the adjective and noun "provide equally new or equally important information". In (5), the fact that Juan's father is a doctor and the fact that he is famous are both new. Adjective-first order is preferred (again, only for unmarked predicates) when the adjective alone provides new or important information. However, Schachter & Otanes claim that these preferences do not hold and that order is "not significant" (122) when the phrase is not an unmarked predicate, as in (6).

(6)   Magara   ang   {pula-ng baro, baro-ng   pula}
      attractive ANG   {red-LINK dress, dress-LINK red}
      'The red dress is attractive'

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 factors of interest, we first obtained words' stress patterns manually from a paper dictionary (English 1986), along with whether they end in a glottal stop. We then coded, partly automatically and partly by hand, the following factors: 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, further grouped into [n], other nasal consonant, non-nasal consonant, or vowel; 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 should take; which one of the words is numeral or one of Schachter and Otanes's "limiters" (121, 141), which include terms that refer to quantity (*marami* 'many', *bahagi* 'part of') and terms that refer to order (*huli* 'last').

3 CONDITIONING FACTORS

Though adjective and 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. Schachter & Otanes, for example, focus largely on syntactic and semantic factors that preference one order over another. Donohue (2007), in contrast, suggests that phonological factors (e.g., weight) condition ordering choice. Donohue further notes that the less-frequent adjective-second order is used in expressing particular contrasts on the adjectives. Even with this range of hypothesized conditioning factors, both Schachter & Otanes and Donohue acknowledge that adjective-noun ordering variation cannot be systematically accounted for by a single predictor alone.

In this section, we introduce the factors that we examined, with their motivations and descriptive statistical results—including factors that in the multivariate model in §4 fail to have a reliable effect. In each case, we provide an example that follows the proposed trend to illustrate the proposal concretely, but these individual examples do not in themselves address whether the trend holds. Furthermore, in this section we examine only whether each trend individually holds in the data overall, not whether it holds once we control for confounding factors—and such controls can cut both ways, showing that an apparent trend was illusory or that an apparent non-trend really does hold, as long as all else is equal. In §4 we present a full multivariate model, and in §0 we return to the implications of some of these factors.

3.1 GENERAL PREDICTORS

3.1.1 Basic "default" order preference

As mentioned above, adjective-first order seems to be the default, with about 90% of tokens showing this order. In our statistical model, this overall preference will be captured by the intercept. In (7), we plot the number of tokens with each order. The difference is highly significant (χ^2=98089.34, df=1, p<0.0001).

(7)   Total number of tokens with each order
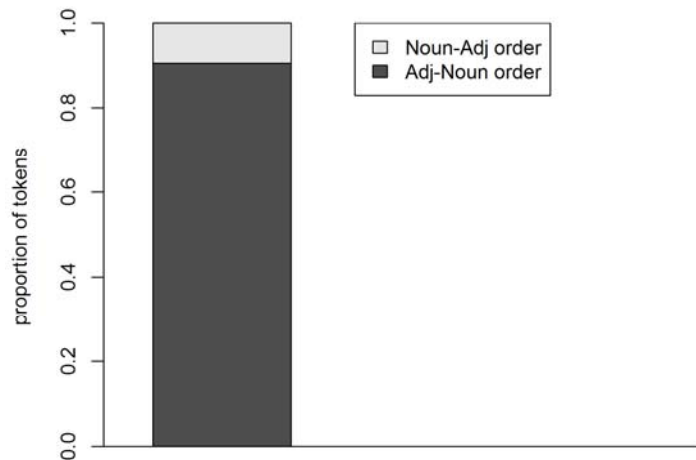
3.1.2   Linker preference

Another overall trend was noted by Schachter & Otanes (1972:122): in a given noun/adjective pair, if one order requires the *-ng* linker and the other requires the *na* linker, then the order that results in the *-ng* linker is preferred. From our corpus, we give two examples in (8) that follow this trend.

(8)   a.   áso-**ng**   ulól   *is more frequent than*   b.   ulól   **na**   áso
               dog-LINK   mad   mad   LINK   dog
               'mad dog'

        c.   bágo-**ng** títser   *is more frequent than*   d.   títser   **na**   bágo
               new-LINK   teacher   teacher LINK   new
               'new teacher'

The plot in (9) shows token counts broken down for each combination of potential linkers. Cases like (8a, b) are plotted as (9a), and cases like (8c, d) are plotted as (9d). Adjective-first order is always preferred, but the preference is much weaker in (9a), where the default prenominal adjective order produces the *na* linker. On the other hand, the preference is slightly stronger than usual in (9d), where the non-default order produces the dispreferred *na* allomorph. The skewing between (9a) and (9d) is significant (χ^2=7161.71, *df*=1, *p*<0.0001).

6

(9)   Token counts by linker type
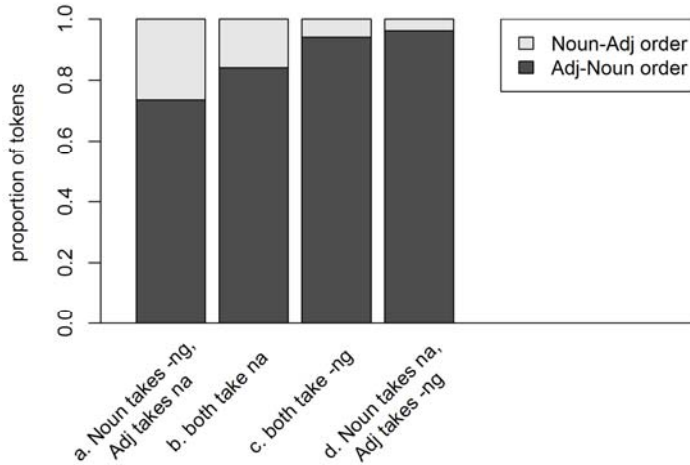
Even though the preference for the *–ng* allomorph may not be itself phonologically-driven—that is, it may be an arbitrary property of the linker—it does *depend* on phonology. To exercise a preference for *–ng*, it is necessary to know the adjective and noun's surface phonological forms.

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, syllable structure optimization, phonological faithfulness.

3.2.1   Weight

Cross-linguistically, a common predictor of constituent ordering has been the "weight"—or length—of a constituent: it has been observed that heavier constituents tend to occur at the peripheries of phrases (Behaghel 1909; Quirk et al. 1985; Hawkins 1994; Wasow 2002 a.o.). English, for example, features the "Heavy-Last Principle," wherein heavier constituents are more likely to come later in phrases. Example (10) illustrates an extreme case from spoken American English, in which a heavy possessor noun phrase occurs after a shorter possessum noun phrase (10a). The alternative order of the heavy noun phrase preceding the shorter one borders on ungrammatical (10b) (example from Shih et al. 2015).

(10)   a.   [the attitude]$_{NP}$ of [people who are really into classical music and feel that if it's not seventy-five years old, it hasn't stood the test of time]$_{NP}$

        b. $^{???}$[people who are really into classical music and feel that if it's not seventy-five years old, it hasn't stood the test of time]$_{NP}$'s [attitude]$_{NP}$

In verb-final languages such as Japanese, on the other hand, the weight effect has been observed to be heavy-first one (Hawkins 1994; Yamashita and Chang 2001 a.o.). One explanation for the

7

underlying cause of heavy-to-periphery phenomena is that moving heavier constituents to the edges aids cognitive processing.[1] In right-branching languages like English, then, 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).

Tagalog word order is generally verb-initial; thus, like English, if a weight preference exists, we would expect a heavy-last tendency. Schachter & Otanes (1972:123) note that relative clause order in Tagalog, which is also variable, has a "tendency to prefer the order head-linker-modifier when the modifying phrase is long." Their example is given in (11). The relative clause, *nilúto mo*, is longer than the head noun, *pagkáin*, and thus prefers to occur second (11a) rather than first (11b):

(11)   a.   ang pagkái-ng nilúto   mo
                  DET food-LINK   cooked you
                  'the food you cooked'

        b.   ang nilúto   mo-ng   pagkáin
                  DET cooked you-LINK   food
                  'the food you cooked'

To operationalize heaviness, two measures each for every noun and adjective in the dataset were coded: the number of phonological segments in a word, and the number of vowel letters in a word, as a close proxy for the number of syllables.[2] Given cross-linguistic patterns of weight, we expect that increasing length of a noun (e.g., *kapangyaríhan*, 'power') will lead to greater likelihood of prenominal, adjective-noun order, as illustrated by

(12)   a.   dakíla-ng   kapangyaríhan
                  great-LINK power
                  'great power'

        b.   kapangyaríha-ng dakíla?
                  power-LINK   great
                  'great power'

Likewise, we expect that the increasing length of an adjective (e.g., *pansamantalá*, 'temporary') will lead to greater likelihood of postnominal, noun-adjective order (13a) over the prenominal alternative (13b):

(13)   a.   lúpo-ng   pansamantalá
                  committee-LINK   temporary
                  'temporary committee'

        b.   pansamantalá-ng lúpon

---

[1] See e.g., (Szmrecsányi 2004; Shih and Grafmiller 2011) for discussion of competing explanations and measures of weight.

[2] A sequence of two orthographic vowel letters in Tagalog is two separate syllables. Diphthongs are written with *y* or *w*: *iw, ay, aw, uy, iy, oy, ow*.

temporary-LINK   committee
'temporary committee'

The figures in (14) demonstrate the probabilities of noun-adjective order in our dataset by the length of adjectives and nouns. The pattern for the increasing length of nouns is as the Heavy-Last Principle predicts: longer nouns are significantly more likely to occur in second position (W = 1058969663, p < 0.0001). For adjectives, however, the data demonstrates a pattern opposite of what is expected under the Heavy-Last Principle. Figure (14) shows that, as the number of vowels in an adjective increases, adjective-first order is increasingly more likely (W = 942912703, p < 0.0001).

(14)   a.   Probability of Noun-Adjective order (versus Adjective-Noun order) by number of vowels in the Noun
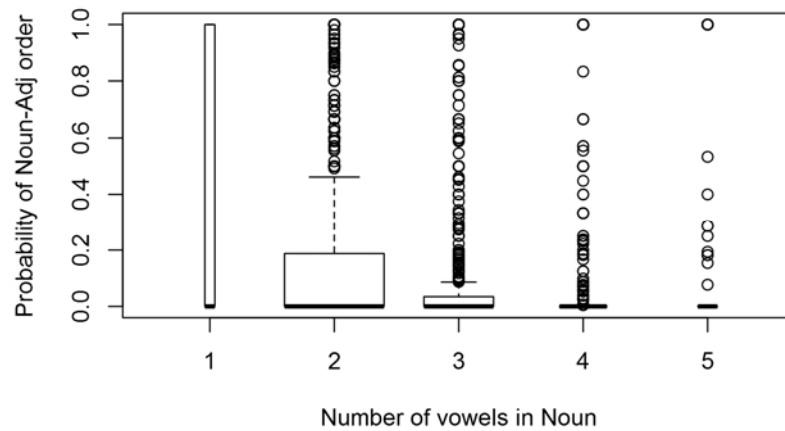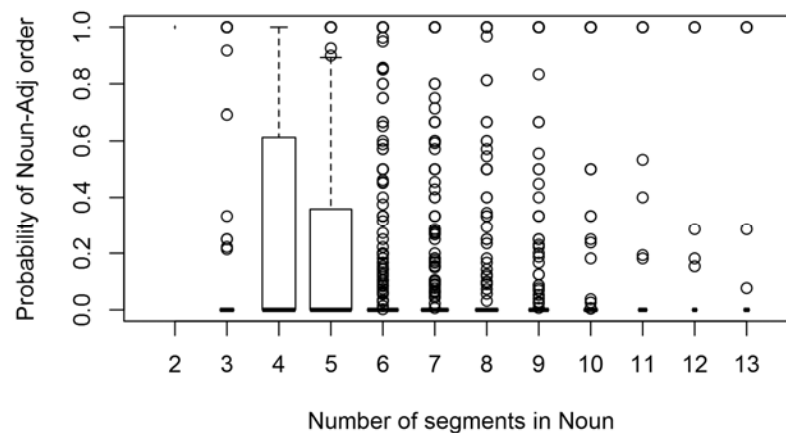

b.   Probability of of Noun-Adjective order (versus Adjective-Noun order) by number of vowels in the Adjective
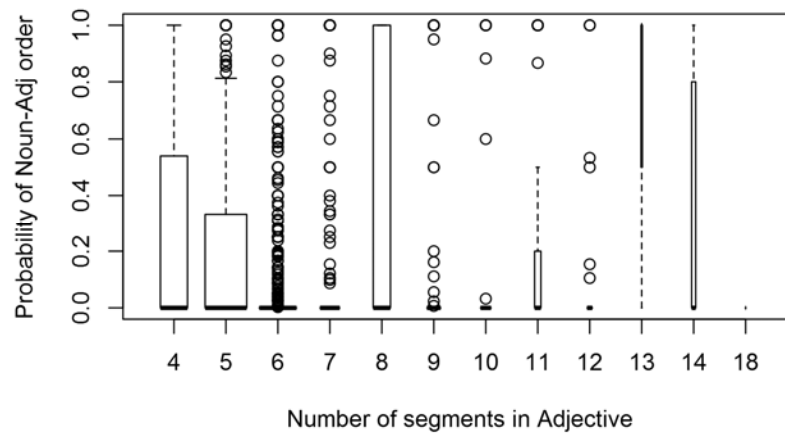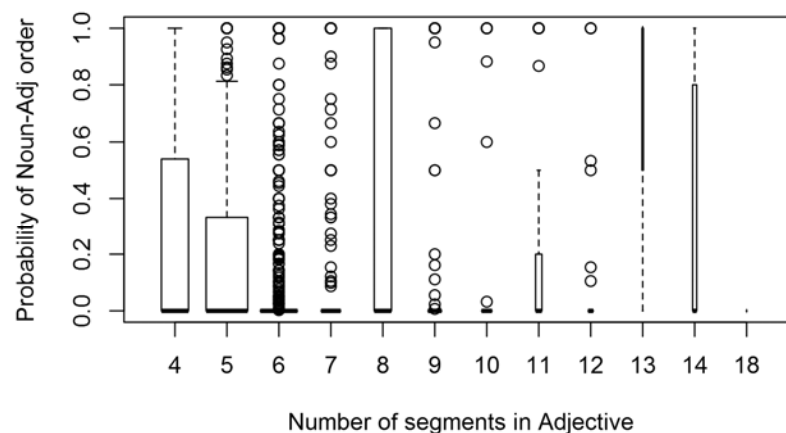

Segment length measures in our data, illustrated in (15), show similar trends to vowel length measures. As nouns increase in length, the likelihood of prenominal ordering increases (15a) (W = 1102856871, p < 0.0001). For adjective length, we find that increases in the number of seg-

9

ments in adjectives correspond to increases in prenominal ordering, with the long adjectives preceding nouns (15b) ($W = 765499276$, $p < 0.0001$). As stated above, that adjective length increases the likelihood of adjective-first order runs counter to the expectations of the Heavy-Last Principle. We will return to this unexpected effect in the presentation and discussion of the multivariate model (§§4–0). We use segment length as the measure of weight in our reported multivariate model (§4) because it allows for the possibility that consonants—both in onsets and in codas—can contribute to weight effects (for in-depth discussion of onset weight, see e.g., Topintzi 2006; Ryan 2011; Ryan 2014; for discussion of coda weight, see e.g., Hayes 1995 and references therein). We tested the same models with vowel length and found no general differences between overall effects, though we leave for future research a direct comparison of weight as measured by segments versus syllables (see e.g., Shih and Grafmiller 2011).

(15) a. Probability of Noun-Adjective order (versus Adjective-Noun order) by number of segments in the Noun

Number of segments in Noun

b. Probability of Noun-Adjective order (versus Adjective-Noun order) by number of segments in the Adjective

Number of segments in Adjective

We also examined here whether the total length of an adjective and noun pair correlates with the likelihood of pre- or postnominal adjective order: does having ever-longer nouns paired

with ever-longer adjectives make adjective-noun (16a) or noun-adjective (16b) more frequent and preferred?

(16) a. pangunáhi-ng katotohánan
          basic-LINK   fact
          'basic facts'

      b. katotohána-ng pangunáhin
          fact-LINK   basic
          'basic facts'

One prediction is that longer constituents that contain long adjectives and nouns may boost the preference for the default prenominal adjective order. In our data, we find that the probability of noun-first order decreases as combined length, measured both in segments (17a) and in syllables (17b), increases. That is, greater total length significantly favors the predominant adjective-noun order (see §3.1.1) ($W = 933301366$, $p < 0.0001$).

(17) a. Probability of Noun-Adjective order (versus Adjective-Noun order) by total number of vowels in the Adjective and Noun pair

b. Probability of Noun-Adjective order (versus Adjective-Noun order) by total number of segments in the Adjective and Noun pair
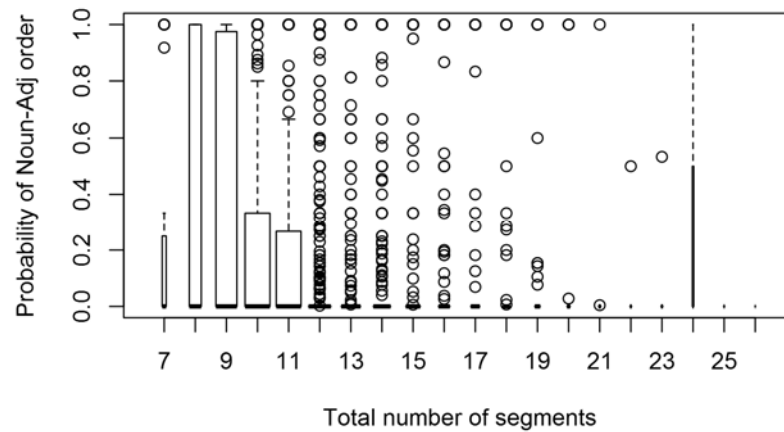
3.2.2 Phonotactics

It is common for phonologies to obey the Obligatory Contour Principle (henceforth, OCP), avoiding sequences of adjacent identical or similar elements (Goldsmith 1976; amongst 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 (Menn and MacWhinney 1984; Zwicky 1987; Hinrichs and Szmrecsányi 2007; Shih et al. 2015; a.o.): e.g., *the rooms of the house*, versus *the hou**se's** rooms*.

To take a clitic-ordering example, in Warlmanpa (Pama-Nyugan, Australia), the reflexive marker /-nyanu/ generally follows person/number clitics (18a). But, when a sequence of adjacent identical nasals would be formed, *-nyanu* instead precedes the clitic (18b) (Wolf 2008:228).

(18) a. -na-nyanu
          1P-REFL

      b. -nyanu-n   *-**n-n**yanu
          REFL-2P

3.2.2.1 Nasal OCP

In Tagalog adjective and noun ordering, the OCP is potentially relevant for nasal and velar segments. For nasals, if either of the words in a pair begins or ends with a nasal, there is a possibility of producing a sequence of immediately adjacent nasals, or a sequence of nasal onsets. For example, in (19), the less frequent order in our data (19b) has a nasal-nasal sequence [m…n], because the word placed first ends in a nasal and takes the *na* linker. Similarly, in (20), the less frequent order in our data (20b) has a nasal-nasal sequence [ŋ…n] (orthographically, *ng…n*), because the word placed first has the *–ng* linker and the latter word begins with [n].

(19)   nasal-C + *na* could be penalized

  a.   pelúka-ng itim   *versus*   b.   iti**m** **n**a   pelúka
        wig-LINK black   black LINK wig
        'black wig'

(20)   –*ng* + nasal-C could be penalized

  a.   naturál na   prodúkto   *versus*   b.   prodúkto-**ng** **n**aturál
        natural LINK   product   product-LINK natural
        'natural product'

For nasals, we also investigated here the possibility of non-local onset similarity avoidance. The interaction of non-local, similar consonants has been demonstrated to have phonological consequences (e.g., harmony; dissimilation) (e.g., McCarthy 1979; Hansson 2001; Rose and Walker 2004). In the Tagalog word order case, a long-distance avoidance would privilege (21a) over (21b), which has, for example, a *na* linker followed by a nasal-initial element.

(21)   *na* + nasal-C could be penalized

  a.   manggá-ng diláw   *versus*   b.   diláw **n**a   **m**anggá
        mango-LINK yellow   yellow LINK mango
        'yellow mango'

Looking at the phonotactics of Tagalog roots, a sequence of two nasal consonants is avoided. In a corpus of 4,294 native, non-reduplicated disyllabic roots (from English 1986), 1,257 have a medial cluster. Of those, 659 (52%) have a nasal as the first consonant, and 66 (5%) have a nasal as the second consonant. We therefore expect 35 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 (19b) and (20b).

By the same reasoning, Tagalog phonotactics supports penalizing the consecutive nasal onsets in (21b). There are 198 roots whose first onset is a nasal, and 565 whose second is a nasal, so we would expect 89 roots whose onsets are both nasal. However, there are only 26.

In our adjective and noun data, we do find that a nasal consonant followed by *na*, as in (19b, *itím na pelúka*), is disfavored. As can be seen in (22), noun-adjective order is much more common than usual in case (22d), 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. Noun-adjective order is also even less common than usual in the case of (22a), when the noun ends in a nasal and takes the linker *na*. The difference between (a) and (d) in (22) is highly significant (χ^2=89.03, df=1, *p*<0.0001).

(22)   Avoidance of nasal C plus *na*
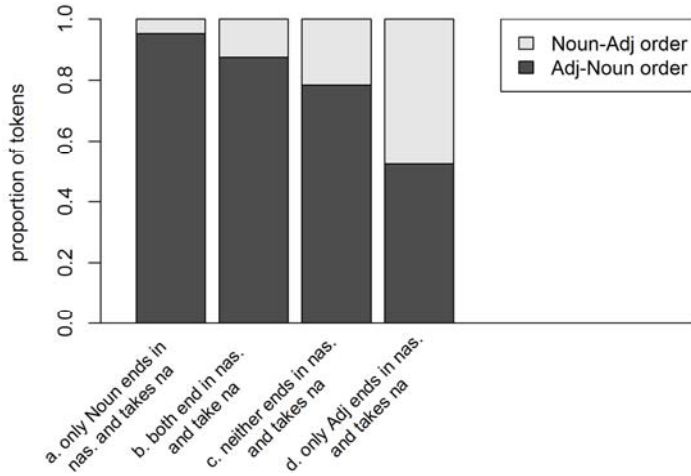
A second form of repeated nasal sequence is linker *–ng* followed by a nasal consonant, producing an immediate nasal-nasal sequence (20b, *prodúkto-**ng naturál***), and linker *na* followed by a nasal consonant, producing successive nasal onsets (21b, *diláw **na manggá***). Comparisons parallel to those above runs into some probable confounds, namely the overall preference for the order that calls for *–ng*, and the ordering preferences *ma-* prefixed adjectives—see §3.3.2). Therefore, we simply compare how often words that follow each allomorph of the linker are nasal-initial:

(23)   Avoidance of nasal C plus *na*
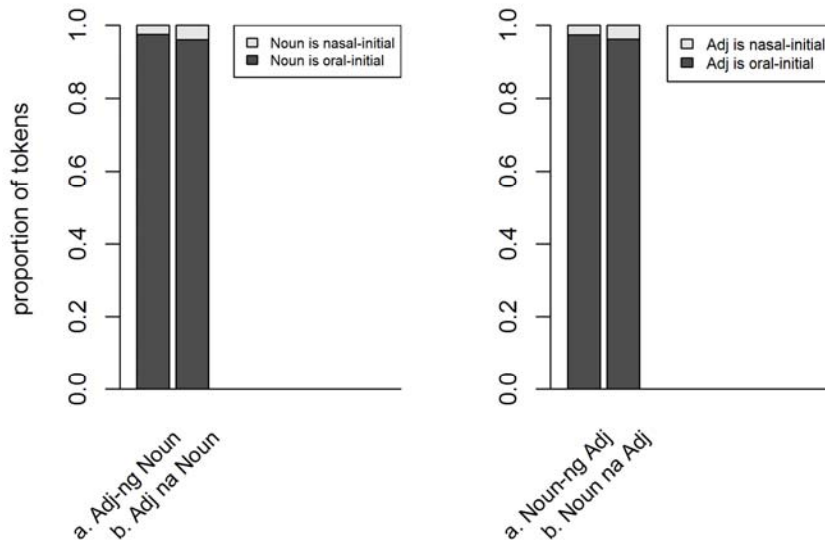
Nouns that follow the *–ng* form of the linker are significantly less likely to be nasal-initial (2.6%) than nouns following the *na* linker (4.0%) (χ^2=163.2991, df=1, *p*<0.0001). Likewise, adjectives (excluding those that begin with *ma-*) are more less often nasal when following *-ng* (2.6%) than when following *na* linker (3.9%) (χ^2=89.0509, df=1, *p*<0.0001). It seems that the local phonotactic effect at the boundary of the adjective and noun sequence is more likely at work than longer-distance similarity avoidance in nasal onset sequences.

3.2.2.2 Velar OCP

The second type of potential OCP violation involves velar segments. If one word takes the *–ng* linker and the other begins with a velar consonant, there is the potential of creating a sequence of two velars:

(24)   *–ng* + velar-C could be penalized

    a.   késo-ng   putí   *versus*   b.   puti-**ng**   **k**éso
           cheese-LINK   white   white-LINK   cheese
           'white cheese'

Phonotactically in Tagalog roots, the picture is different for velar-velar sequences, as compared to nasal-nasal sequences. In 497 clusters, the first consonant is velar, and in 165 clusters, the second consonant is velar. We therefore expect 65 velar-velar clusters if consonants combined freely, and in fact, there are 152—many more than expected. The great majority of the clusters are [ŋg] and [ŋk]. The phonotactics of roots do not therefore support a penalty for velar-velar sequences.

Our adjective and noun ordering data show no penalty for velar-velar sequences either. We would expect case (25d) to show increased noun-first order, but it actually shows less noun-first order than (25a). Again, we exclude confounding *ma*-prefixed adjectives for the illustration here, since these adjectives never begin with a velar. It is curious that (25b) shows such a high rate of adjective-noun order; we have no explanation for this, but note that there are far fewer tokens in that category than in the other three (352, versus 4454, 69402, and 8365).

(25)   No avoidance of *ng* plus velar C
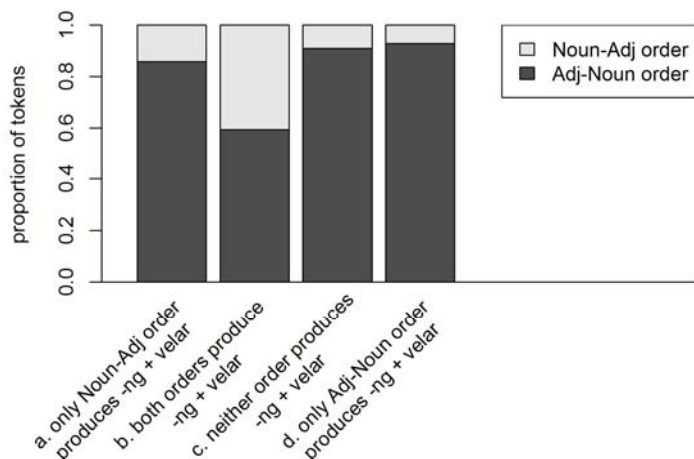
3.2.3 Contextual markedness: *NÇ

Cross-linguistically, sequences of a nasal consonant followed by a voiceless obstruent (NÇ clusters) can trigger a number of phonological repairs. such as deletion, assimilation, and fusion (Pater 1996; Pater 2001). Within Tagalog words, the *NÇ constraint (Pater 1996; Pater 2001; Hayes and Stivers 1996) 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 (26a), but a voiced obstruent usually does not (26b):

(26) a. /ma-paŋ-**k**amkám/ → [ma-pa-**ŋ**amkám] 'rapacious'
      b. /paŋ-**d**ing/ → [pa**n**-**d**inig] 'sense of hearing'

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

(27) a. péra-ng   nakalaán   *versus*   b. nakalaá-**ng**   **p**éra
          money-LINK   dedicated   dedicated-LINK   money
          'dedicated money'

Because *NÇ is confounded with the linker allomorph, which we showed in §3.1.2 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 non-nasal consonants. The overall data shows the opposite of what we would expect: adjective-first order appears to be more common than usual when it produces a NÇ cluster (28d). Our multivariate model will show instead that when other factors are controlled for, the *NÇ effect is in the predicted direction.

(28) Expected effect of *NÇ does not hold in raw data
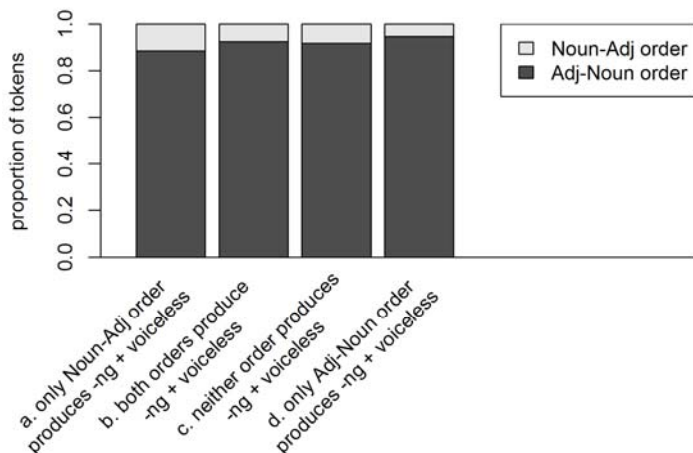
3.2.4   Phonological and morphological alignment

It is common for morpheme boundaries to prefer to coincide with syllable boundaries. For example, McCarthy & 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 this alignment constraint to be at work in Axininca Campa, Lardil, Hebrew, Bedouin Arabic, and Kamaiurá. Alignment between syllable boundaries and higher-level boundaries can affect the ordering of syllables in these cases as well. For example, the choice between output [na.ta] and *[ta.na] in Axininca Campa depends on leftward alignment (ALIGN-L).

   Alignment is relevant to Tagalog word order because if one word takes the linker *–ng* and the other begins with a vowel (29b), then there will be either misalignment, with the word final velar nasal serving as an onset (e.g., [pu.lá.ŋit.lóg]), or insertion of a glottal stop (e.g., [pu.láŋ.ʔit.lóg]), or a syllabification that violates both NOCODA and ONSET (e.g., [pu.láŋ.it.lóg]) (see (Tranel and Del Gobbo 2002) on *C.V). Putting the vowel-initial word first (29a) will avoid the alignment problem if the item is preceded by a vowel-final word, or by a large prosodic boundary.

(29)   a.   **i**tlóg   na   pulá   *versus*   b.   pulá-**ng**   **i**tlóg
            egg   LINK   red   red-LINK   egg
            'red [brined] egg'

In (30), we see that adjective-second order is even less common than usual when it produces *–ng* + vowel (a). The difference between cases (a) and (d), where prenominal adjective order produces the undesirable, misaligned sequence, is highly significant (χ^2=714.86, df=1, *p*<0.0001).

(30)   Morpheme and syllable alignment
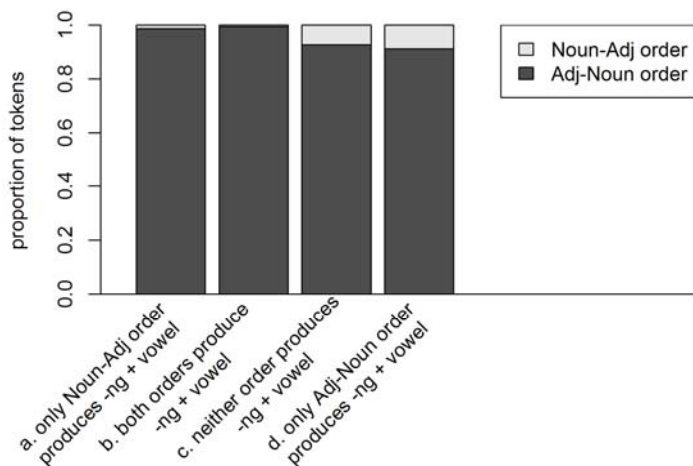
3.2.5   Syllable structure optimization

In the ordering of coordinated noun pairs in English, two vowels in a row tend to be avoided (Wright et al. 2005; Benor and Levy 2006 a.o.):

(31) a. John and Yoko *versus* b. Yok**o a**nd John

In general, hiatus avoidance has widespread effects in phonology, and the constraint ONSET is a fundamental one in Optimality Theory (Prince and 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. The sequence *na vowel* will either violate ONSET or require a [ʔ] to be epenthesized, whereas the opposite order might avoid the problem entirely:

(32) a. espesyál na bágay *versus* b. bágay na ?**e**spesyál
        special LINK thing   thing LINK special
        'special thing'

As the plot in (33) shows, hiatus avoidance is a trend in our data. Adjective-first order is less common than usual in (33d), where that order produces the environment that forces misalignment, hiatus, or epenthesis, and the reverse order avoids the misalignment. The difference between (33a) and (d) is significant (χ^2=13.5, df=1, p=0.0002).

(33) Hiatus avoidance
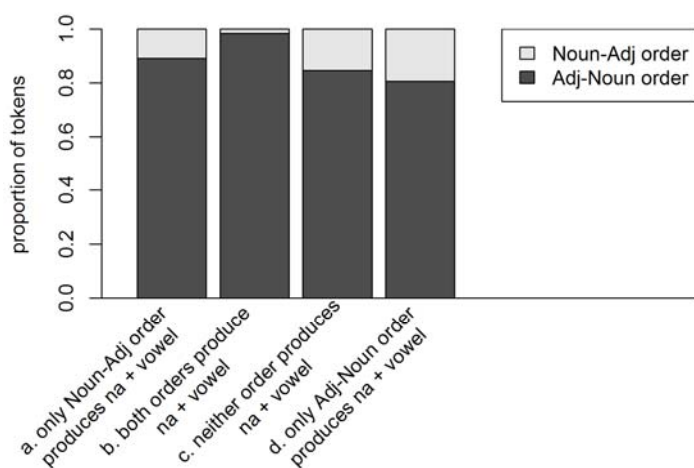
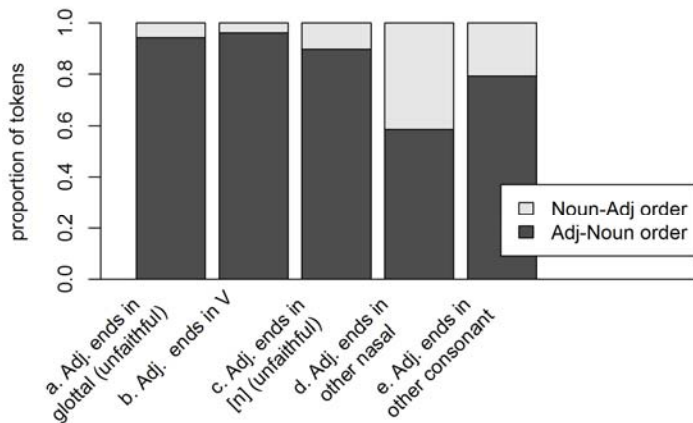*3.2.6 Phonological faithfulness*

Recall that when the first word in an adjective and noun pair ends with a [ʔ] or [n], attaching the linker particle causes the final consonant to be replaced by [ŋ]. This ought to incur a faithfulness penalty, and so might be avoided. In (34), we see that nouns that end with a [ʔ] (34a) have a higher rate of coming first (and so incurring a faithfulness penalty) than nouns that end in a vowel (34b). This is the opposite effect from what we expect. Nouns that end in [n] (34c) are also slightly more likely to come first in ordering than nouns that begin with other consonants (34d, e), again the opposite of what is expected under a faithfulness hypothesis.

(34)   No overall faithfulness effect for nouns
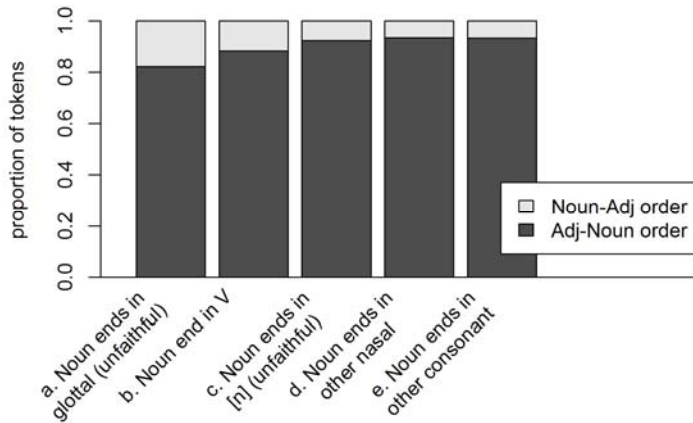
Dividing up adjectives in the same way, we see slightly less prenominal adjective order when the adjective ends in a [ʔ] than in a vowel (35a, b), which is the direction expected under phonological faithfulness (χ^2=179.98, df=1, p<0.0001). We do not, however, see the expected lowered rate of adjective-first order in (35c), as compared to (35d, e).

(35)   No overall faithfulness effect for adjectives

Consistent with the overall results just described, phonological faithfulness does not present to be reliable in the multivariate analysis presented below, and is not included in our final model.

3.2.7   A note on prosody

Previous work on other languages has shown that prosodic rhythmicity can be achieved using alternative word order choices. In English, stress clash—a sequence of adjacent stressed sylla-bles—can sometimes be avoided through the use of another word order choice (see also Tem-perley 2009). For example, clash avoidance has been argued to be the reason that *a*-initial adjectives, such as *aware* and *asleep*, have undergone a diachronic change so that they are now not allowed to be prenominal, as in (36b, d) (Schlüter 2005). Instead, postnominal order (36a, c) is used, to avoid the potential stress clash.

(36) a.   σ́ σ σ σ σ σ́   b.   σ σ́ σ́ σ

            the **pér**son who was a**wáre**   [?]the a**wáre pér**son

            σ́ σ σ σ σ σ́   σ σ́ σ́ σ

c.   the **pér**son who was a**sléep**   d. ^{??}the a**sléep pér**son

A stress lapse, consisting of a sequence of adjacent unstressed syllables, has also been shown to be avoidable through word order (e.g., McDonald et al. 1993; Wright et al. 2005; Benor and Levy 2006; Shih et al. 2015; Shih 2014). In the genitive alternation for American English, for example, Shih et al. (2015) demonstrate that the genitive construction that avoids long stress lapses (37a) will be preferred over the alternative that incurs long lapses (37b).

(37) a.   σ́ **σ**   σ́ σ σ   b.   σ́ **σ** σ σ σ σ́ σ

            the sýstem's bénefits   the bénefits of the sýstem

Given the rhythmicity patterns from other studies of word order variation, we might expect that Tagalog, a stress-based language, would have similar prosodic conditions on adjective and noun order. In Tagalog, primary stress is contrastive, and can be final, penultimate, or in some loans, antepenultimate.[3] Stressed non-final vowels are notably longer than unstressed ones. As mentioned above in §2, stress is not indicated in Tagalog orthography, and we manually coded it using a paper dictionary (i.e., English 1986). Secondary stress is not fully understood in Tagalog (French 1988; French 1991 a.o.), and it is usually not included in dictionary entries, so we did not annotate our data for it at this point.

It has been argued that the avoidance of stress clash and stress lapse are at work in Tagalog phonology, at least in the assignment of secondary stress (French 1988:71ff). Assuming that the same principles affect adjective and noun order, we could expect to see variation based on stress clash and lapse avoidance, as is seen in other languages. If one word ends in a stressed syllable and takes the linker –ng and the other word begins with a stressed syllable, stress clash could result in one order but not the other (38a versus b). Alternatively, 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 violation (38d).

(38) a.   σ́ σ   σ σ́   b.   σ σ́   σ́ σ

           báya-ng   sawíʔ   sa**wí-ng**   **bá**yan

           country-LINK   unfortunate   unfortunate-LINK   country

           'unfortunate country/people

c.   σ σ σ́ σ   σ́ σ   d.   σ́ **σ** σ   σ **σ** σ́

       espesyál na   áraw   **á**raw **na**   **espe**syál

       special   LINK   day   day   LINK   special

       'special day'

Because we only have primary stress information at this point, we can preliminarily examine disyllabic and trisyllabic words, to avoid possible noise from unknown secondary stresses. The da-

---

[3] In Schachter & Otanes' (1972) analysis, it is length that is contrastive, and the phonology places stress on long vowels.

ta presented in (39) and (40) are further restricted to just those words whose stress patterns are entirely known,[4] and the rare antepenultimate stress pattern is also excluded.

In a preliminary exploration, what we see is a slight trend in the overall data towards less noun-first order when that order alone would produce a stress clash: 7.6% for (39a) versus 9.8% when neither order produces clash (39b), and 9.1% for when only adjective-first order produces clash (39c) (χ^2=18.02, df=1, p<0.0001). However, there is only a slight difference, and in the un-expected direction, between (39b) and (39c). (Because we are restricting the data to multisyllabic words, there are no cases where both orders would produce a stress clash.)

(39)   Stress clash
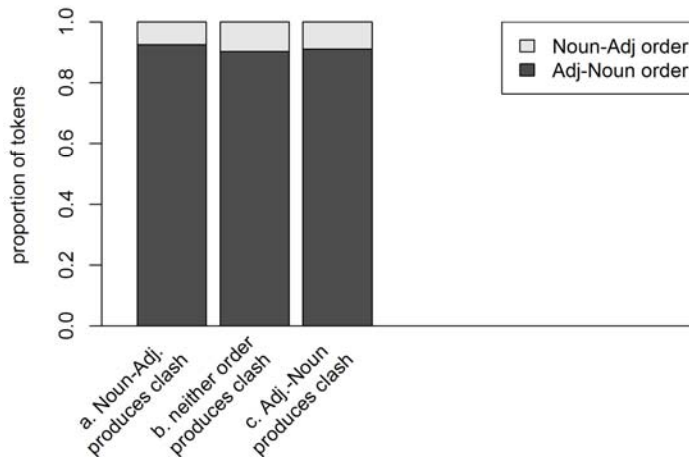
The sharpest comparison for stress lapse comes from items where one order produces a long lapse (i.e., 4 unstressed syllables in a row, incurred by the combination of unstressed *na* linker flanked by sequences of unstressed syllables) and the other does not. Those data are plotted in (40). The rate of adjective-second order is lowest when the order would produce a long lapse (40a) and the rate of adjective-first order is lowest when that order would produce a long lapse (40c). The difference between those two cases is highly significant (χ^2=293.85, df=1, p<0.0001). (There are no cases where both orders produce a long lapse.)

---

[4] Sometimes the existence of homographs with different stress patterns render some or all of the token's stress unknown.

(40)   Stress lapse
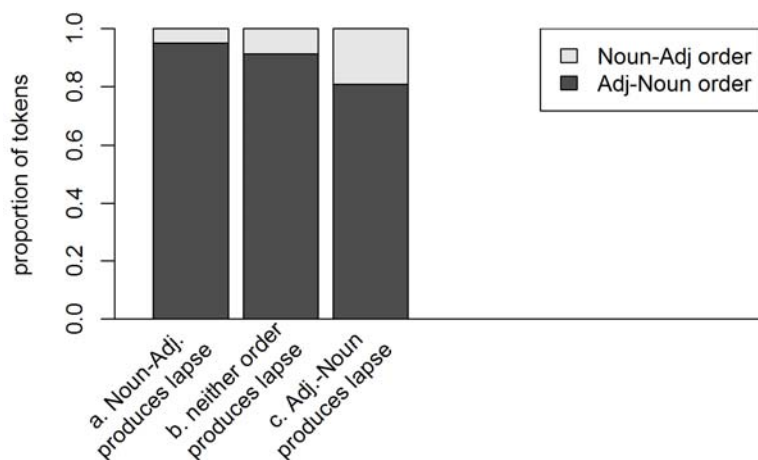
The preliminary exploration presented here demonstrates that rhythmicity in clash and lapse avoidance may play a role in word order variation, but, at this time, not enough is understood about the rhythmic well-formedness preferences of Tagalog on the whole to be able to informatively investigate these prosodic factors as influences in word order variation. We will reserve these prosodic factors and their potential relationship to word order variation in Tagalog for future investigation.

3.3   NON-PHONOLOGICAL PREDICTORS

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


3.3.1   Quantifier adjectives

Schachter & Otanes note that prenominal adjective order is more likely to occur when an 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" (1972:141). An example of a variable quantifier adjective from our corpus is given in (41).

(41)   **lahát** na   táo   ~   táo-ng   **lahát**
       all   LINK person   person-LINK all
       'all people'

We coded adjectives for their membership in limiter and numeral classes using a list taken from Schachter & Otanes (provided in Appendix A). To simplify the analysis, the limiter and numeral classes were collapsed into a binary quantifier coding in the following investigation: adjectives are categorized as either quantifiers or not.

22

Following Schachter & Otanes, the prediction is that adjective-noun pairs in which the adjective is of the quantifier set will appear in adjective-first order and disprefer adjective-second order, all else being equal. In our corpus, as plotted in (42), we observe that quantifiers are significantly more likely to come first than non-quantifiers (χ^2=1659.09, df=1, p<0.0001).

(42)   Quantifiers
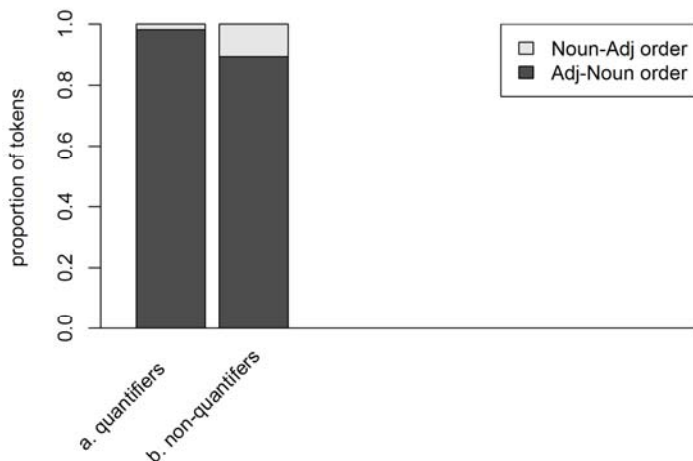
3.3.2   *Adjective prototypicality*

Items that are prototypical are ones that are more conceptually central to membership in a given category: for example, for American English speakers, an apple is generally considered to be a more prototypical example of a fruit than a loquat (or a lemon) is. Prototypical items have been shown to be more easily accessed in language processing, which results in earlier ordering in a phrase than non-prototypical items: e.g., *an apple and a loquat* is more likely than *a loquat and an apple* (Kelly et al. 1986; Onishi et al. 2008; a.o.).

In Tagalog, adjectives are often formed using the *ma-* prefix: for example, *ma-baho* 'malodorous' from *baho* 'bad smell.' Thus, *ma*-initial adjectives are prototypical adjectives in phonological form. 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 of segments *ma* but do not actually have the affix *ma–*: e.g., *mahal* 'dear') prefer prenominal order. In 45% of our tokens, the adjective begins with the string *ma*.

In (43), we exclude quantifiers; we also exclude vowel-initial adjectives, because, as we have seen, they show a strong tendency to occur first. Comparing only non-quantifier *ma*-initial adjectives to other consonant-initial non-quantifier adjectives, the difference is very small (87.9% vs. 87.3%), but significant (χ^2=9.33, df=1, p=0.002). In the multivariate model below, we will see that when other factors are controlled for the effect is significant.

(43)   Prototypical versus non-prototypical adjectives
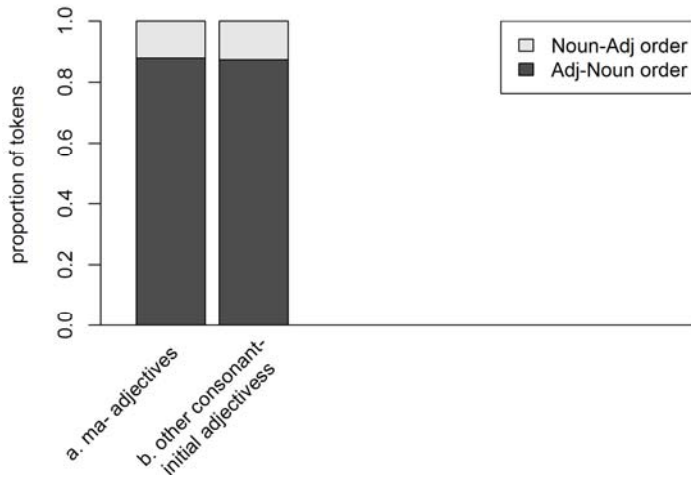
3.3.3   Frequency

From a processing point of view, more frequent items are hypothesized to be more readily accessible (e.g., Griffin and Bock 1998, and references therein); therefore, more frequent items are expected to be more likely to occur earlier than less frequent items. We control for frequency in our study by including individual adjective and individual noun frequencies, as well as adjective and noun pair frequencies.[5]

Our expectation is that the more frequent word will occur first. The effects of individual adjective and individual noun frequencies are illustrated in the frequency density plots in (44a) and (44b), respectively. In (44a), pairs of adjectives and nouns that consistently appear in the adjective-noun order are shifted slightly higher on the adjective frequency scale, as compared to pairs that consistently appear in noun-adjective order. Similarly in (44b), pairs with consistent adjective-noun order have more frequent nouns.

(44)   a.   Densities for adjective log frequencies for consistently noun-adjective pairs versus consistently adjective-noun pairs


---
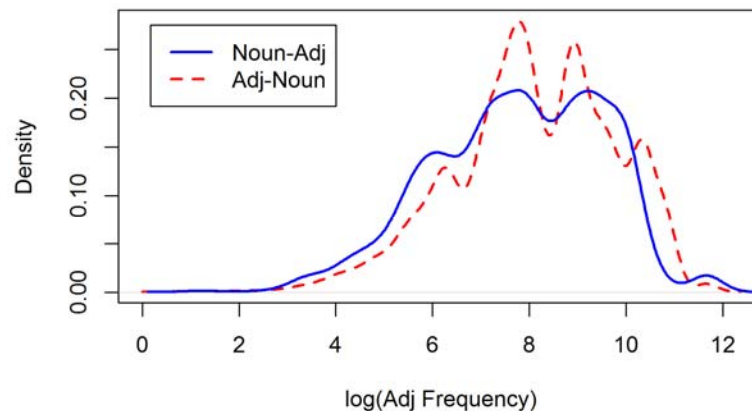
b.   Densities for noun log frequencies for consistently noun-adjective pairs versus consistently adjective-noun pairs
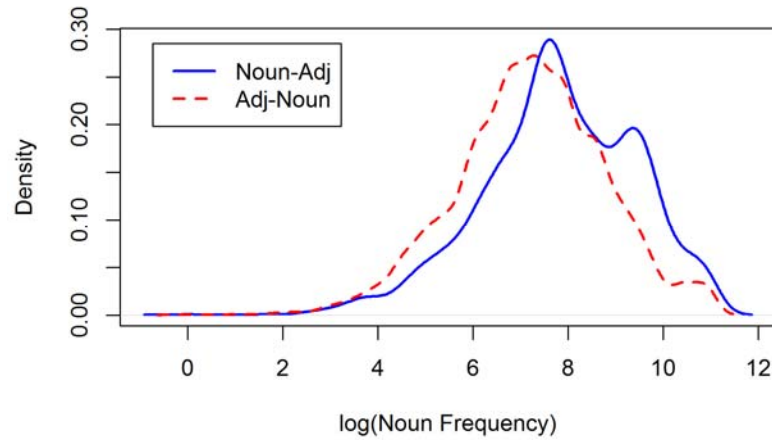
The plot in (45) compares pair frequencies for pairs that appear consistently in each order. While very little effect is noticeable in the plot, the multivariate regression model that is presented in the following section (§4) finds that, after residualizing by individual noun and individual adjective frequencies, higher pair frequency promotes noun-adjective order (when the adjective takes *–ng*).

(45)   Densities for adjective and noun pair log frequencies for consistently noun-adjective pairs versus consistently adjective-noun pairs
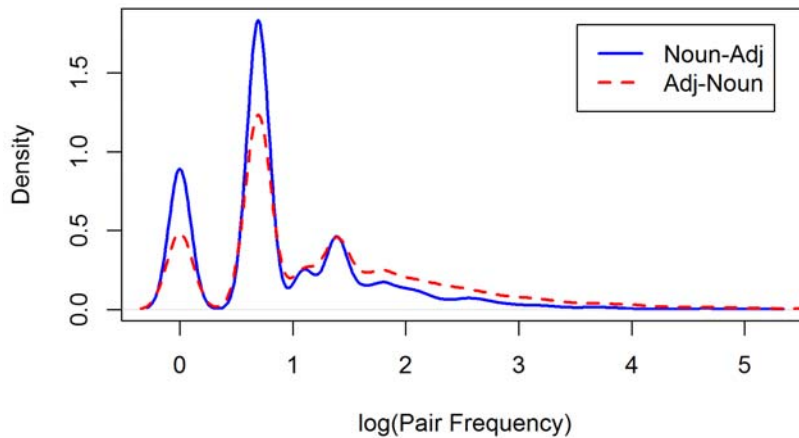
3.4   SUMMARY OF PREDICTORS

The predictors we will test using multivariate regression modeling in the following section are summarized in (46).

(46) Predictors to be tested

General conditioning predictors	
Default order preference:	Adjective-Noun preferred
Linker:	Adjective (*–ng* preferred)
	Noun (*–ng* preferred)
	–ng preferred over *na*
Phonological predictors	
Weight:	Adj length
	Noun length
	Total length
*[nasal][nasal] OCP:	Avoid nasal-final Adj + *na/–ng*
	Avoid nasal-final Noun + *na/–ng*
	Avoid *–ng* + nasal-initial Adj
	Avoid *–ng* + nasal-initial Noun
*[velar][velar] OCP:	Avoid *–ng* + velar-initial Adj
	Avoid *–ng* + velar-initial Noun
Long-distance OCP:	Avoid *na* + nasal-initial Adj
	Avoid *na* + nasal-initial Noun
Contextual markedness:	Avoid *–ng* + Ç-initial Adj
	Avoid *–ng* + Ç-initial Noun
Alignment:	Avoid *–ng* + V-initial Adj
	Avoid *–ng* + V-initial Noun
Hiatus avoidance:	Avoid *na* + V-initial Adj
	Avoid *na* + V-initial Noun
Phonological faithfulness:	Avoid ʔ-final Adj
	Avoid ʔ-final Noun
Non-phonological predictors	
Quantifier Adjective	
Prototypicality	
Frequency	Noun
	Adjective
	Pair

Predictors that we did not examine, for practical reasons, but that we expect could have an effect on adjective and noun order variation include the following: being part of a predicate, givenness, focus, and animacy. As mentioned in §2, an adjective and noun phrase used as a predicate has different ordering properties, according to Schachter & Otanes. If either the noun or the adjective represents information already given in the discourse, 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 and Szmrecsányi 2007; Shih et al. 2015). Similarly, if one of the words is under focus of some kind and the other is not, order could well be affected. Finally, animacy has been found to be important in word order considerations (e.g., Rosenbach 2005), and animate versus inanimate nouns might show different order preferences. The influence of these factors is saved for future investigation.

4   RESULTS

In this section, we present the results of a multivariate model of adjective-noun word order patterns in the Tagalog web corpus data. We used mixed effects logistic regression modeling to examine the effects and reliability of each conditioning factor introduced in §3) in predicting word order. Modeling was done using the `glmer()` function of the `lme4` R package (Bates et al. 2013).[6] 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 helps to mitigate some potential data multicollinearity, and facilitates 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 potential variation between individual lexical items in their propensity for one or the other adjective and noun order. A random intercept for each unique noun and adjective 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.[7]

The final model was fit using stepwise backwards 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 $\in [-1.5, 1.5]$): factors removed were also checked against the descriptive patterns discussed in §3. 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 liberal 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 non-default, postnominal noun-adjective order (e.g., positive regression coefficient indicates increased probability of noun-first order; negative regression coefficient indicates decreased probability of noun-first order). In the logistic model, predicted probability of noun-adjective order is estimated as a linear combination of the intercept and factor weights, as shown in (47).

(47) $p(y) = \dfrac{1}{1 + e^{-(x' \cdot \beta)}}$,

where y = noun-adjective order;
β = vector of fitted coefficients (i.e., weights); and
x' = transposed vector of predictor variables, including 1 for the intercept.

The results of the final multivariate model are given in (48).

---

[6] 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.

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

(48)   Modeling estimates

Factor		Estimate	Std. Err	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						
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	***
	(nasal-initial A)	0.2240	0.3864	0.580	0.5621	
	(Ç-initial A)	-0.1785	0.3127	-0.571	0.5681	
	(Ç-initial N)	0.1795	0.1405	1.277	0.2016	
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.	*NÇ: linker (N)=*na* * Ç-initial A	0.1992	0.1014	1.966	0.0493	*
xii.	*NÇ: linker (A)=*na* * Ç-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	.
Non-phonological 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	*
Non-phonological interactions						
xviii.	linker (A)= *na* * A/N pair freq (resid)	-0.6540	0.2594	-2.521	0.0117	*
Random effects		Variance		Std. Dev.		
	Noun (intercept)	4.5813		2.1404		
	Adjective (intercept)	6.1983		2.4896		
	N = 14,591	AIC_C = 31075.2				
	Nouns = 1708	Marginal R^2 = 0.1440; Conditional R^2 = 0.7998				
	Adjectives = 587	κ = 4.6794				

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

The final model presented above 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 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 §3.4, 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).

    The following tables in (49)–(51) summarize the hypothesized conditioning factors that were tested and indicates which factors were reliable predictors of Tagalog adjective and noun word order in our dataset. Note that the regression model results in (48) show the independent variables as they were coded for modeling; in (49)–(51) and in subsequent discussion, the condi-

tioning predictors we tested and found reliable are indexed to their corresponding factors in the model above with small Roman numerals.

(49)	*General conditioning predictors tested*		*Index*	*Effect?*
	Default order preference (Adj-Noun)		(i)	✓
	Linker:	Adjective (*–ng* preferred)	(ii)	✓
		Noun (*–ng* preferred)	(iii)	✓
		–ng preferred over *na*	(viii)	✓

(50)	*Phonological predictors tested*		*Index*	*Effect?*
	Weight:	Adj length	(iv)	✓
		Noun length	(v)	✓
		Total length	(ix)	✓
	*[nasal][nasal] OCP:	Avoid nasal-final Adj + *na*/*–ng*	(vi)	✓
		Avoid nasal-final Noun + *na*/*–ng*		✗
		Avoid *–ng* + nasal-initial Adj	(x)	✓
		Avoid *–ng* + nasal-initial Noun		✗
	*[velar][velar] OCP:	Avoid *–ng* + velar-initial Adj		✗
		Avoid *–ng* + velar-initial Noun		✗
	Long-distance OCP:	Avoid *na* + nasal-initial Adj		✗
		Avoid *na* + nasal-initial Noun		✗
	Contextual markedness:	Avoid *–ng* + Ç-initial Adj	(xi)	✓
		Avoid *–ng* + Ç-initial Noun	(xii)	✓
	Alignment:	Avoid *–ng* + V-initial Adj	(vii)	✓
		Avoid *–ng* + V-initial Noun		✗
	Hiatus avoidance:	Avoid *na* + V-initial Adj	(xiii)	✓
		Avoid *na* + V-initial Noun		✗
	Phonological faithfulness:	Avoid ʔ-final Adj		✗
		Avoid ʔ-final Noun		✗

(51)	*Non-phonological predictors tested*		*Index*	*Effect?*
	Quantifier Adjective		(xiv)	✓
	Prototypicality		(xv)	✓
	Frequency	Noun	(xvi)	✓
		Adjective		✗
		Pair	(xvii)	✓

Our results reveal that both phonological and non-phonological 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., phonological faithfulness) is a reliable predictor of word order. 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—no matter how reasonable—can elicit a positive result from the data modeling.

The model results in (48) 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 conditioning factor that the model showed to be reliable using a drop-

one, nested models test of increase in second order Akaike Information Criterion (AIC$_c$) (Burnham and Anderson 2002; Burnham and Anderson 2004; a.o.). Second order AIC$_c$ is calculated based on the log-likelihood and number of parameters of a model:

(52) $\quad AIC_c = -2log\left(\mathcal{L}(\hat{\beta}|D)\right) + 2K + \frac{2K(K+1)}{n-K-1},$

> 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$_c$ is a measure of how much information a model captures, given the evidence provided. It is calculated using model maximum log likelihood, but also crucially takes into account the number of parameters used in model estimation, thereby rewarding parsimonious models that capture more information with the fewest number of 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 (53). Following Burnham & 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 an equal to or greater than 10-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 non-phonological predictors. For example, if the predictor for nasal-final adjectives (48:vi, *[nas][nas] OCP) is removed, the resulting model would be far worse than a model for which the predictor for quantifier adjectives (48:xiv) were removed. Individual predictor results are discussed in the following sections.

(53)   AIC$_c$ results of predictor importance
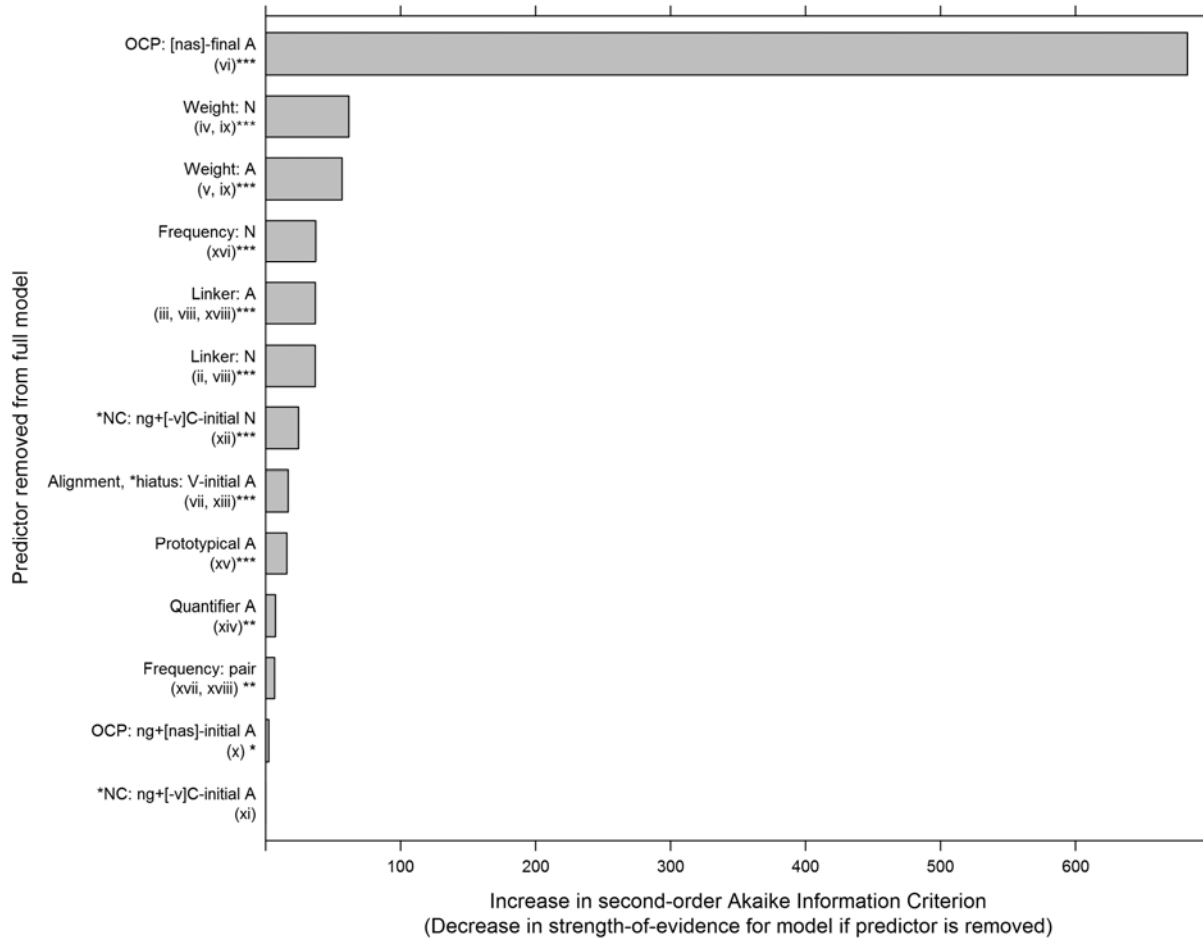
4.1   GENERAL PREDICTORS RESULTS

4.1.1   *Basic "default" order preference*

The regression model results confirm that adjective-first order is significantly more common than adjective-second order. This is captured by the negative intercept term of the regression model (48:i): when all else is held constant at the mean, there is a 91.98% probability of adjective-first order.[8] We find that the prenominal order behaves like a "default" order in the language. In the discussion of other factors, we will see that the prenominal adjective word order is often favored: the effect sizes are often larger for adjective-first order, and adjective-first order is often preferred when increased phonological or processing difficulty is present.

---

[8] $p(\text{Adjective– Noun Order}) = 1 - \left(\frac{1}{1+e^{-(-2.4398)}}\right) = 1 - \text{logit}^{-1}(-2.4398) = 0.9198$. (For description of probability calculations, see (47).)

4.1.2 Linker preference

As noted by Schachter & Otanes (1972) (see §3.1.2), the model demonstrates that there is 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 (54). The *y*-axis represents increasing probability of noun-first order, versus adjective-first order.

(54)
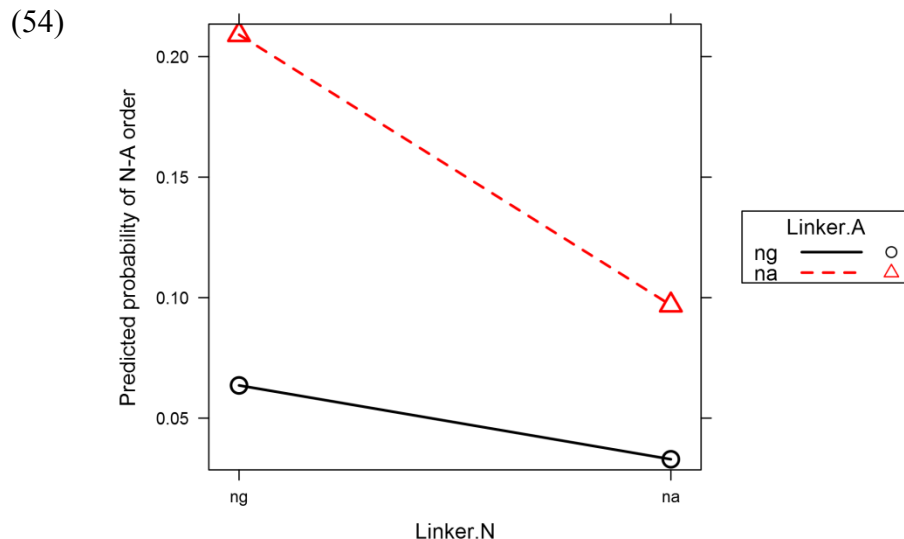
As shown in (54), there is an overall preference for the *–ng* linker form. When the noun takes the *–ng* linker, the predicted probability of noun-first order is greater than when the noun takes the *na* linker (represented on the *x*-axis). When the adjective takes the *–ng* linker, adjective-first order is more likely than when the adjective takes the *na* linker, as shown by the contrast between the lower, solid line and the upper, dashed line in (54), respectively.

   We find a significant interaction effect between the linkers of the adjectives and nouns. Given a noun and adjective pair, if the noun takes the *–ng* linker and the adjective takes the *na* linker (e.g., *áso-ng ulól* 'dog-LINK mad' versus *ulól na áso* 'mad LINK dog'), then the probability of noun-first word order is 11.22% greater than if both noun and adjective took *na* linkers (i.e., if bout noun and adjective are equal with respect to linker surface form)[9]: this effect is represented in (54) by the steeper slope of dashed line (Adjective linker = *na*) as compared to the solid line. 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.

---

[9] $= 100 \times \left(\mathrm{logit}^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 y_0 + \beta_3 x_1 y_0) - \mathrm{logit}^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 y_1 + \beta_3 x_1 y_1) \right),$

where $\beta_0 = -2.4398$ (intercept);
$\beta_1 = 1.2793$ (Adjective linker coefficient);
$x_0 = -0.3990131$ (Adj. linker = *ng*);
$x_1 = 0.6009869$ (Adj. linker = *na*);
$\beta_2 = -0.7744$ (Noun linker coefficient);
$y_0 = -0.37763$ (Noun linker = *ng*);
$y_1 = 0.62237$ (Noun linker = *na*); and
$\beta_3 = -0.2121$ (linker interaction coefficient).

The model results show that whichever order promotes a *–ng* linker will be the one that is most likely to surface. The surface form of both potential linkers—the linker of the adjective and the linker of the noun—compete in the model. Linker preference is shown in (53) above to be an important contributing predictor in the model. We find in the regression model a stronger size of effect for the form of the adjective linker (β=1.2793) than for the noun linker (β=-0.7744); however, the importance the adjective linker form does not significantly outrank the importance of the noun linker in predictive adequacy of the model (ΔAIC_c=0.07). Both linker preferences are shown to be nearly equivalent in how much they contribute to influencing adjective and noun order.

Our results of linker preference suggests that knowledge of the surface, phonologically-conditioned allomorph form of the linker must be present at the time of word order choice. That we find an interaction effect further points to some amount of competition between the linkers of nouns and of adjectives: both the surface linker allomorph of the noun and the surface linker allomorph of the corresponding adjective must be known in order to undergo comparisons in which *–ng* wins out over *na*.


4.2 PHONOLOGICAL RESULTS

Amongst the phonological predictors of variable adjective and noun word order tested, we find that the reliable and important contributing factors to predicting word order include weight, adjacent nasal avoidance, contextual markedness (*NÇ̥), 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 and noun order included the following: adjacent velar avoidance, long distance segmental phonotactics, and phonological faithfulness. These results concur with the basic descriptive statistics presented in §3.2, which also found that Tagalog adjective-noun word order variation exhibits no differences for adjacent velar avoidance, long distance segmental phonotactics, and phonological faithfulness. The reliable factors found in the model are discussed in turn below.
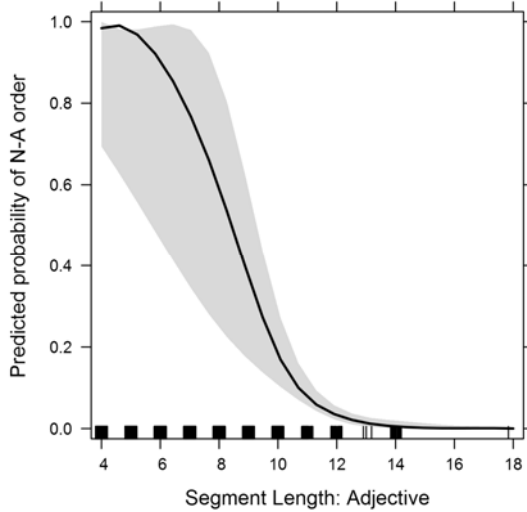

4.2.1 Weight

As with the basic descriptive statistics in §3.2.1, the regression model in (48) finds significant effects of weight, measured here by segment length, in predicting adjective and noun word order variation. The model results for the main effects of noun and adjective segment lengths are illustrated in (55) and (56), respectively.

(55)
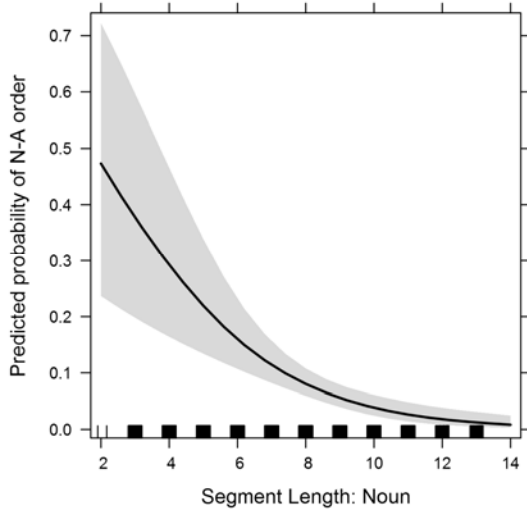
(56)

Under the hypothesis of Heavy-Last shift, we expect that longer words shift towards 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, 6 segment noun to a 9 or 10 segment long noun results in a 1.2% decrease in noun-adjective order probability.[10] Longer nouns favor adjective-first order, as expected under the Heavy Last Principle.

For adjectives, we find an opposite effect. Running counter to heavy-last expectations, the model reveals that longer adjectives prefer adjective-first order. Holding noun length constant, an increase from an average length, 6 segment adjective to a 9 segment long adjective also results in a decrease the probability of noun-adjective order, by 2.4%. Prenominal adjective-first order seems to be preferred whenever the length of either noun or adjective constituents increase. The result that increasing adjective length also leads to prenominal adjective-first order is sur-

---

[10] $= 100 \times \left(\text{logit}^{-1}(\beta_0 + \beta_1 x_1) - \text{logit}^{-1}(\beta_0 + \beta_1 x_0) \right)$,
where $\beta_0 = -2.4398$ (intercept);
$\beta_1 = -0.1798$ (Noun length coefficient);
$x_0 = 0$ (Noun length mean); and
$x_1 = 1$ (2 standard deviations from the Noun length mean).

prising, given a heavy-last expectation. What we see instead is that longer adjectives lead to a greater probability of the "default" adjective-noun order. The behavior of pairs with long nouns also prefers the default order, but in the case of nouns, the default order aligns with the heavier noun being in final position.

The model demonstrates an interaction between the length of the adjective and noun in a given pair, illustrated in (57).

(57)
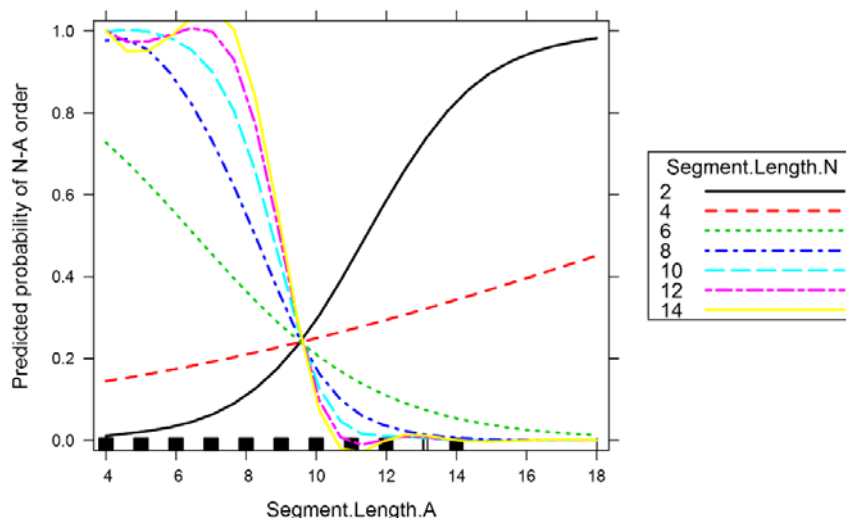
The graph in (57) reveals that, when the noun is very short (e.g., 2–4 segments long), then a heavy-last effect for longer adjectives does emerge. But when the noun is longer (i.e., more than 4 segments), a default order preference asserts itself: combinations of long nouns and long adjectives are more likely to occur in adjective-first order.

Usual processing-based explanations of weight effects have assumed that peripheral weight effects facilitate the balancing of 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 the use of default order removes the cognitive uncertainty of adjective-noun word order variation. When processing difficulty is compounded by the use of both ever-longer nouns with ever-longer adjectives, reverting to a default order lightens the processing load by enforcing some certainty in signaling an adjective and noun complex. Though such an explanation of default order preference is consistent with uncertainty minimization effects in psycholinguistics (e.g., Frank and 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 to variable adjective and noun word order, second only to one predictor of adjacent nasal avoidance.

4.2.2 Phonotactics

Two phonotactic effects were found to be reliable and important to predicting adjective and noun order in Tagalog: both were Obligatory Contour Principle effects, avoiding sequences of adjacent nasal segments (*[nasal][nasal]). Other OCP, phonotactic constraints that were tested—for

instance, the avoidance of adjacent velar sequences—were not found to be reliable in predicting word order variation.

The model reveals a main effect of nasal OCP when an adjective has a final nasal, illustrated in (58). When the adjective is nasal-final, adjective-first order is strongly dispreferred, 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').

(58)

*[nas][nas]: Nasal-final A
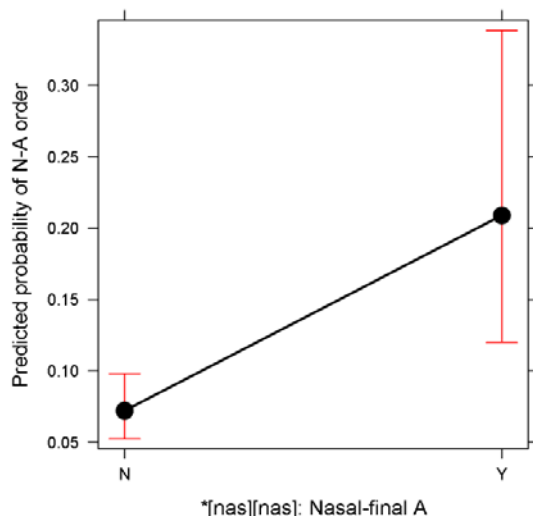
Postnominal adjective order is predicted to be 13.69% more likely when an adjective is nasal-final, compared to a non-nasal-final adjective. This OCP avoidance is furthermore shown to be the most important factor in contributing to word order prediction: the fit of a model without this specific nasal OCP factor drops drastically (ΔAIC_C=682.92). In importance, avoidance of a nasal-final adjective followed by the linker particle outranks even psycholinguistic (e.g., frequency, weight) and semantic factors (e.g., semantic class of adjective), which are known to have strong effects on constituent ordering cross-linguistically that normally trump phonological factors (see e.g., Hinrichs and Szmrecsányi 2007; Grafmiller 2014; Shih et al. 2015 on genitive construction choice in English).

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 (59).

(59)
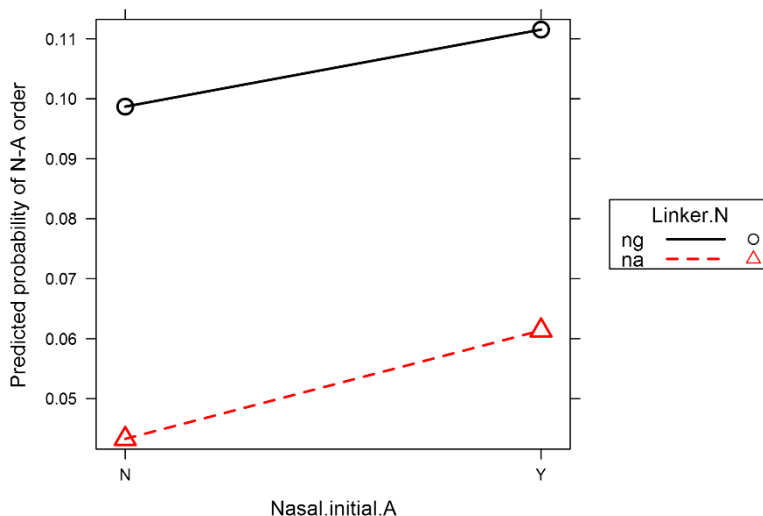
When an adjective is nasal-initial and the noun would take an *–ng* linker, we find an avoidance of adjective-second order as compared to when the noun would take a *na* linker. Graphically, the effect is shown by the solid line in (59), representing a noun with a *–ng* linker, having a shallower slope than the dotted line, which represents a noun with a *na* linker. As shown by the AIC test in (53), the avoidance of a noun with the *–ng* linker followed by a nasal-initial adjective contributes significantly less in the explanatory power of the model when compared to the avoidance of nasal sequences in an adjective-first situation (i.e., factor (48vi); shown in (58)). We will see that many of the reliable and important effects found in modeling adjective-noun ordering variation will be strongest when they govern well-formedness preferences in the default, adjective-first order.

4.2.3   Contextual markedness: *NC̥

Contextual markedness effects, as noted in §3.2.3, were confirmed in the multivariate model results: adjective and 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 (60) 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', vs. *NC̥-violating *nakalaá-**ng** péra*), then noun-adjective order is more likely than when the noun begins with a voiced consonant. In comparison, there is no effect of consonant voicing when the adjective takes the *na* linker (e.g., neither *diyós na túnay* 'God-LINK true' or *túnay na diyós* violates *NC̥). The plot in (61) demonstrates a similar pattern for cases in which the adjective begins with a voiceless consonant and the noun takes the *–ng* linker. Here, adjective-first order is significantly more likely when adjective-second order would produce a marked NC̥ cluster.

(60)
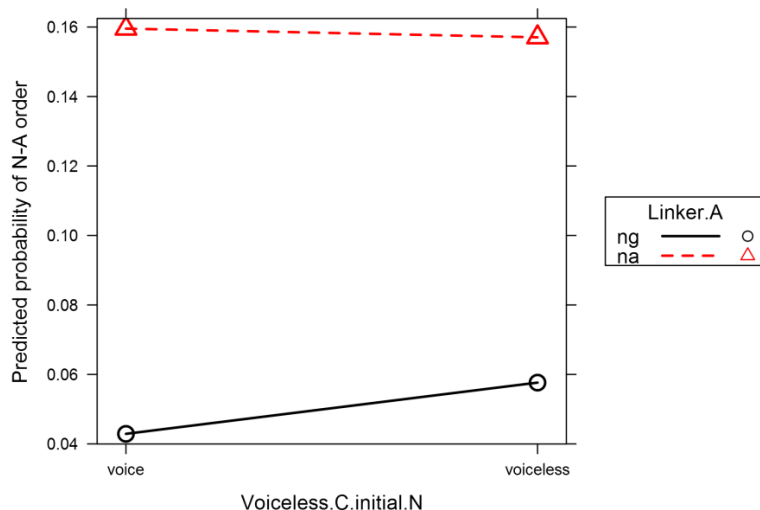
(61)
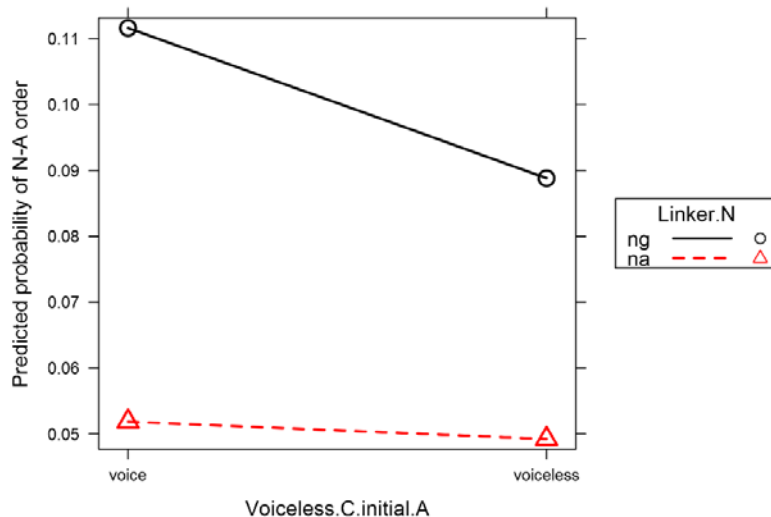
AIC$_c$ comparisons (53) indicate that avoiding NÇ clusters that would be formed via a combination of an adjective followed by a noun is an important contributor of the model: the default word order (adjective+noun) is likely to be flouted when it would result in a marked phonological configuration. For cases of nouns followed by adjectives, the NÇ avoidance effect is reliable but not as important of an explanatory parameter in the model of word order presented here: this AIC$_c$ result suggests that the phonological configuration of the non-default order (noun+adjective) may not be as crucial to considerations of word order choice. We will discuss this difference in default and non-default ordering results further in §0.

4.2.4   Alignment and syllable structure optimization

The effects of alignment and syllable structure optimization are plotted in (62). As the plot demonstrates, noun-adjective order is most likely noun takes the –*ng* linker and is followed by a consonant-initial adjective, which respects both syllable structure and morphophonological 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 i**tlóg*, '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] epethensis (e.g., *bágay **na ˀespesyál***, 'thing LINK special').

(62)
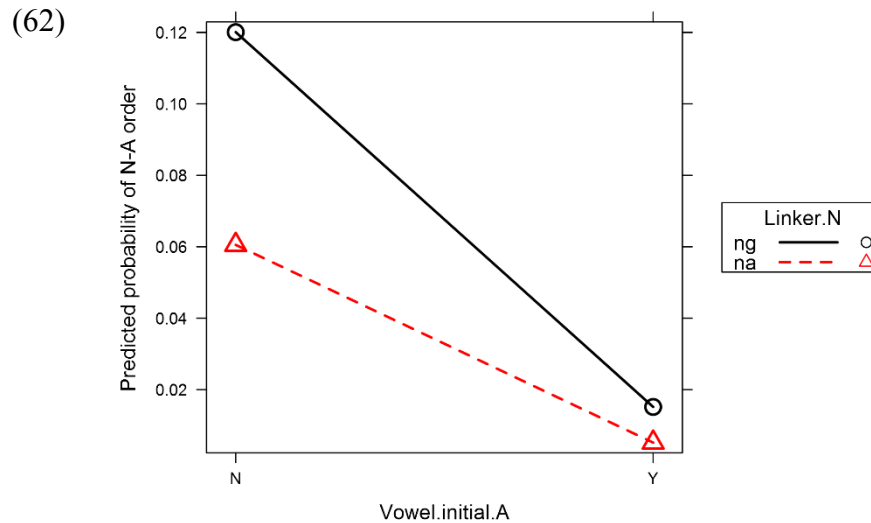
4.3 NON-PHONOLOGICAL RESULTS

4.3.1 Adjective prototypicality

As was expected, adjectives that begin with *ma-*, which mark adjectival prototypicality (see §3.3.2), are significantly biased towards adjective-first order. The effect is illustrated in (63). Prototypicality, despite being a semantic, non-phonological predictor, is shown by the drop-one, AIC_c test to contribute less explanatory importance to the model than many phonological effects.

(63)
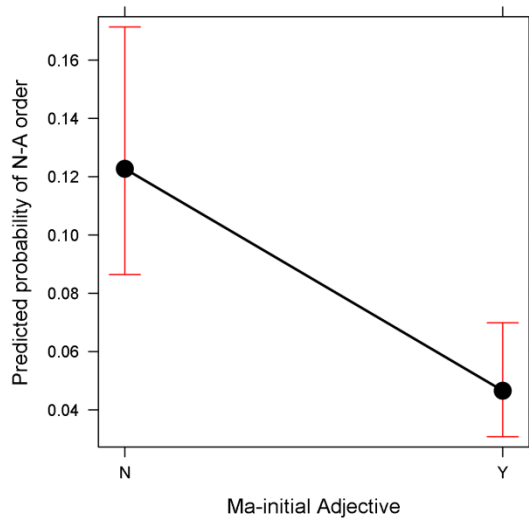
4.3.2 Quantifier adjectives

Illustrated in (64), the model shows that when the adjective belongs to the quantifier class, default adjective-first 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 takes the prototypical phonological *ma*- shape, and if it is a quantifier—then the more likely default adjective-noun order occurs. It is, however, notable that, like prototypicality, the quantifier status of an adjective is not as important of a conditioning factor as many phonological conditioners, including, for example, phonotactic effects like contextual markedness (see AIC_c (53)).

(64)
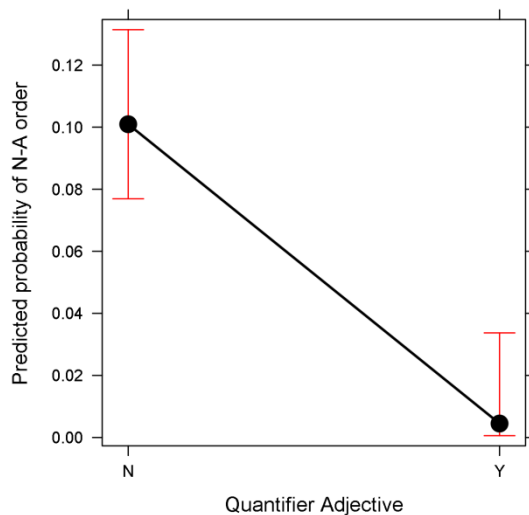
4.3.3 Frequency

As noun frequency increases, the likelihood of noun-first order also increases. An increase from the mean of 1600 by one standard deviation (on the log scale) to 8000 tokens in noun frequency in the corpus results in a 13.29% increase in noun-first order. This result accords with expectations of the role of frequency in lexical access: more frequent items are more easily accessed 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 OCP avoidance and weight.

(65)
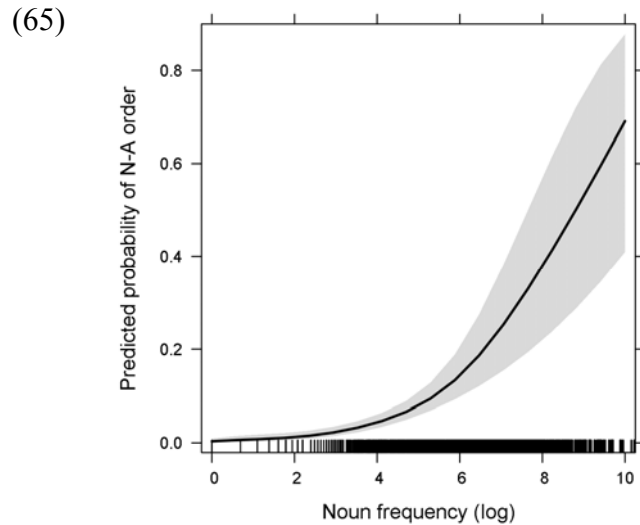
For pair frequency, we find that higher-frequency pairs—after controlling for individual noun and individual adjective frequencies—are slightly more likely to occur in noun-adjective order, if the adjective takes the –*ng* linker, as illustrated in the left plot of (66). There is no effect for cases when the adjective takes *na*, as shown in the right plot.

(66)
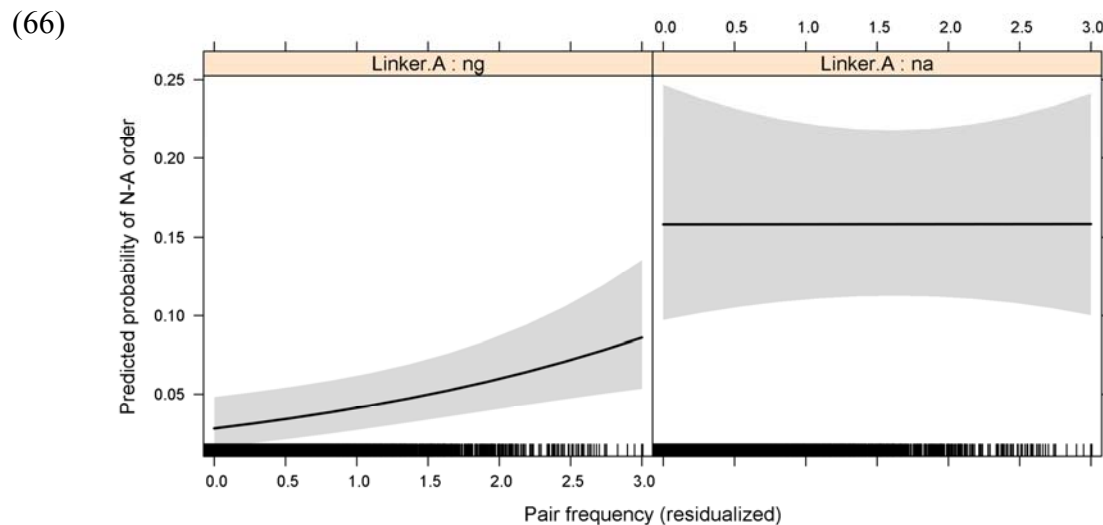
5   Comparison to a speech corpus

We used our dictionary-derived lists of nouns and adjectives to search for adjective/noun pairs (with linker) in the the Philippine Speech Corpus (Guevara et al. 2002), which includes about 5 minutes each of spontaneous speech from 65 talkers, transcribed into normal spelling.

     Noun/adjective data are sparse. We identified 44 tokens (38 types), of which 26 involve a quantifier adjective and 18, a non-quantifier adjective.[11] None of the quantifiers occur second, and of the non-quantifiers, only one token (*báta-ng payát* 'child-LINK thin') occurs in noun-adjective, non-default order. This token demonstrates the preference of the *–ng* linker form to the *na* form: the alternative word order would produce *payát na báta* 'thin LINK child', with the less preferred *na* linker (see discussion in §4.1.2). Similarly as expected by our model, 16 of the 17 other tokens of non-quantifier adjectives that occur in adjective-noun order always demonstrate preference for *–ng* (e.g., observed *bágo-ng gámit* 'new-LINK thing', versus non-observed *gámit na bágo* 'thing LINK new'). The single counterexample (observed *mataás na paaralán* 'high LINK school', versus non-observed *paaraláng mataás* 'school-LINK high') exhibits a prototypical adjective, which, given our model's prediction, biases the pair towards adjective-noun ordering in spite of the *na* linker form (see e.g., §4.3.1).

     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-first order, probably stronger for quantifiers, with adjective-second 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 & Conclusion

The investigation presented here examined how phonology influences word order variation in Tagalog. Our corpus study revealed that, controlling for non-phonological predictors, surface phonological information and phonological well-formedness conditions significantly contribute to predicting variable adjective and noun order, an alternation that had previously been characterized as primarily "free."

     Studies of constituent order variation from other languages have demonstrated similar effects of phonological influence (e.g., McDonald et al. 1993; Wright et al. 2005; Benor and Levy 2006; Mollin 2012; Shih et al. 2015). The existing literature has zeroed in on certain phonological factors that have been shown repeatedly to influence constituent order variation, across a number of different constructions: prosody, and weight (e.g., Zec and Inkelas 1990; Zubizarreta 1998; Anttila et al. 2010; Ehret et al. 2014). Individual studies have also demonstrated effects of adjacent segmental identity avoidance (e.g., adjacent sibilant avoidance in English genitive constructions: Hinrichs and Szmrecsányi 2007; a.o.); long-distance phonotactic preferences (e.g., high to low vowel height ordering in English: *teeny tiny* versus ???*tiny teeny*: Cooper and Ross

---

[11] Looking more broadly, we did come across a few additional examples in the transcripts of adjective-second order (but that were not found by our procedure, because either the noun or the adjective is not in our list), such as *kapatid na nakatatandang* 'brother LINK elder'.

1975; though cf. Benor and Levy 2006[12]); and syllable structure optimization (e.g., hiatus avoidance in English binomial order: Wright et al. 2005; a.o.). To this growing list of possible phonological conditioning factors of word order, our study adds the effects of contextual markedness (e.g., *NÇ) and morphophonological alignment.

One 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? And, importantly for future work in understanding ordering variation, are there uniting principles that can guide more systematic identification of crucial phonological factors from the entire grab bag of possibilities?

One early proposed answer is that phonological influences on constituent order variation should be limited to prosodic conditions, which include weight prosodically-defined), phrasal stress, and intonation (Selkirk 1984; Inkelas 1990; Zec and Inkelas 1990; a.o.). The idea is that higher-level prosodic components interface with (limited amounts of) syntactic information. Lower-level segmental phonology, on the other hand, interfaces solely with prosody, not syntax (see also e.g., Zwicky and Pullum 1986a; 1986b; Levelt 1989; Vogel and Kenesei 1990; Bock and Levelt 1994; Miller et al. 1997; Garrett 2000). 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.

Given that segmental effects have been identified in the current study and elsewhere as influencing constituent ordering, 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. Our working proposal instead is that the phonological conditioning factors that play a role in influencing constituent order variation are the ones that also affect syntagmatic configurations elsewhere in the phonological grammar of the language (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 and 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 for the speaker and language learner from within-word phonological patterns, that is, will tend to generalize into across-word domains in the (probabilistic) grammar of a language.

Such a view, then, allows that word order conditioning factors would come from the prosodic domain *and* from the rest of the phonological system. We expect 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 of adjective and noun combination 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 one that does not avoid the 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 prediction intrinsic to this proposal is that, for the phonological effects identified here to influence adjective and noun word order, the same conditions are likely to be observable

---

[12] For a non-English example, see e.g., (Mortensen 2006:222–223) for vowel height preferences in Jingpho compound coordination.

elsewhere in the phonotactics and morphophonological alternations of Tagalog. Similarly, phonological conditions that have no import in Tagalog phonology are not expected to have as much potential to condition word order variation. Of course, possible phonological effects may still arise from the presence of certain conditions universally present across languages, but given that speakers will have less evidence that these conditions are present in Tagalog, we do not predict that these universal conditions generalize across domains as readily to become phonological conditions word order.

The two local Obligatory Contour Principle (OCP) conditions that we tested in this study—*[nasal][nasal] and *[velar][velar]—illustrate our prediction. As discussed in §3.2.2, adjacent clusters of [nasal] segments are underrepresented in the lexicon, whereas adjacent clusters of [velar] segments are not, with medial [ŋk] and [ŋg] clusters being common. We find the parallel pattern in our study of adjective-noun word order: orderings which result in adjacent nasals are avoided (e.g., *[nasal]-final Adjective + linker), and no reliable effect of velar cluster avoidance is uncovered. 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. Note that the avoidance of identical-place segments does exist cross-linguistically. For example, Arabic verbal roots exhibit an underrepresentation of sequences of dorsal segments as compared to sequences of nasal segments (e.g., OCP-PLACE » OCP-MANNER) (McCarthy 1979; Frisch et al. 2004; a.o.). So, the parallelisms we find between within-word and without-word patterns in Tagalog do not arise from the complete, cross-linguistic absence of [velar][velar] avoidance.

Though it is not a perfect predictor, the expectation that well-formedness constraints operating in other parts of the phonological grammar will, with stronger likelihood, operate in across-word phenomena like constituent ordering seems to be borne out in our study here. 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, the contextual markedness condition avoiding sequences of a nasal segment followed by a voiceless consonant (*NÇ) compels deletion of the non-nasal consonant when nasal-final prefixes and stems combine.

As we find at the adjective-noun boundary, morpheme-syllable alignment also applies within words, where resyllabification does not happen across prefix-stem boundaries. For instance, the combination of a consonant-final prefix with a glottal stop-initial stem fails to trigger deletion of the glottal and resyllabification, even though within a morpheme, consonant+[ʔ] sequences are extremely rare: [mag-] + [ʔalis] 'departure' → [magʔalis] 'remove', *[magalis] (example from Schachter and Otanes 1972:19).

On our approach, opening the door to the possibility that segmental phonology can influence word order does not mean that *any* segmental information can willy nilly influence word order. We predict that only the phonological constraints present cross-linguistically and most especially those that carry weight of 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. Our results from Tagalog adjective and noun ordering fit with this prediction.

Our findings here also raise the question of how adjective-noun word order is chosen in Tagalog, and how phonological conditions interface with this process: at what point in language production and/or the grammatical system do the phonological conditioning factors identified here, as well as other non-phonological factors, influence adjective and noun ordering?

We briefly consider two extreme models of the morphosyntax-phonology interface: [1] both adjective-noun and noun-adjective orders are generated as viable outcomes and then compared for well-formedness on the basis of the conditioning factors, which we call here the *post-morphosyntactic filter model*, or [2] one order is derived from another, more basic order, according to phonological well-formedness considerations, which we call the *interleaved phonology model*. Roughly speaking, in the post-morphosyntactic filter model the phonological information does not intervene in generating (presumably, equally probable) morphosyntactically grammatical options, and filters only after the morphosyntax has been completed (e.g., Embick and Noyer 2001; Pak 2008). In psycholinguistic terms, this would mean that grammatical encoding and phonological encoding are independent processes that experience only unidirectional information flow from the grammatical to phonological modules (e.g., Levelt 1989; Bock and Levelt 1994; Ferreira and Slevc 2007). On the other hand, the interleaved phonology model allows phonological information to participate in the grammatical linearisation process of spell-out. In some accounts of linearisation in Distributed Morphology, for example, this interleaving of linearisation and phonological information has at least been posited for prosodic structure-building (e.g., Tucker 2011), or as multiple cycles through the PHONOLOGICAL FORM branch of the grammatical model (e.g., Embick 2007)—though crucially, to account for the Tagalog data, post-morphophonological, surface segmental information must also be available, since, for example, the surface phonological form of the linker particle plays a role in order preferences. In psycholinguistic terms, the interleaved approach calls for bidirectional information flow between grammatical encoding and phonological encoding (e.g., Vigliocco and Hartsuiker 2002).

Our corpus investigation suggests that the two possible orders of adjectives and nouns in Tagalog are not equal options, as a post-morphosyntactic filter model assumes. First, adjective-first order is overwhelmingly preferred, even once controls for individual words and frequencies are included. This quantitative results aligns with observations in the previous Tagalog literature that one order is more basic—or, *default*—than the other (see discussion in §§3.1.1 and 4.1.1).

Moreover, we find that the conditioning factors with the strongest effect sizes tend to be the ones that regulate phonological well-formedness in the prenominal adjective-noun order—that is, phonological ill-formedness in the *default* order appears to be more strictly penalized. For example, although the surface form of the linker for both adjectives and nouns contributes order choice, the linker form of the adjective is a stronger effect (Adj. linker β=1.2793, vs. Noun linker β=-0.7744). Similarly, nasal-nasal OCP avoidance (see results in §4.2.2) is regulated more strongly for cases in which a prenominal adjective is nasal-final (β=1.224) than for cases where a nasal-initial adjective is post-nominal, following an –*ng* linker (β=0.224). The results also show a greater magnitude of effect of *NÇ in the default adjective-noun order, where a voiceless consonant-initial noun would form an illicit NÇ cluster with an –*ng* linker on the adjective (Adj-Noun β=-0.3299, vs. Noun-Adj β=0.1992, see §4.2.3). One possible exception to this otherwise strong pattern of prenominal order primacy is alignment and hiatus avoidance, for which our model shows an effect for vowel-initial adjectives in noun-adjective order but not noun-initial adjectives in adjective-noun order.

The model of adjective-noun word order choice reported here points to a preference for a default word order, for which there are more and stronger active phonological (and non-phonological) constraints that penalize phonological ill-formedness in the default order. We take these results as evidence for an interleaved phonology model, in which some amount phonological surface information (that has undergone morphophonological operations) feeds back into the process of morphosyntactic linearisation. Of course, it is most likely that language production

and the grammatical system must integrate elements of both extreme views compared here—i.e., post-morphosyntactic filtering and competition as well as interleaved, cyclic phonological feedback. We leave this as an open question for future work, and offer Tagalog adjective and noun word ordering as an example of a type of variable case study that could illuminate crucial interactions between morphophonological alternations and morphosyntactic order in natural language.

REFERENCES

ANTTILA, ARTO; MATTHEW ADAMS; and MICHAEL SPERIOSU. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25.946–981.

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* 43.209–226. doi:10.1007/s10579-009-9081-4.

BARTOŃ, KAMIL. 2013. *Package "MuMIn."* http://cran.r-project.org/web/packages/MuMIn/index.html.

BATES, DOUGLAS; MARTIN MAECHLER; and BEN BOLKER. 2013. *Package "lme4."* http://cran.r-project.org/web/packages/lme4/lme4.pdf.

BEHAGHEL, OTTO. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25.110–142.

BENOR, SARAH BUNIN; and ROGER LEVY. 2006. The Chicken or the Egg? A Probabilistic Analysis of English Binomials. *Language* 82.233–278.

BIBER, DOUGLAS; STIG JOHANSSON; SUSAN CONRAD; and EDWARD FINNEGAN. 1999. *Longman Grammar of Spoken and Written English*. Harlow, England: Pearson Education, Ltd.

BOCK, KATHRYN; and WILLEM LEVELT. 1994. Language Production: Grammatical Encoding. *Handbook of Psycholinguistics*, ed. by M.A. Gernsbacher, 741–779. New York, NY: Academic Press, Inc.

BURNHAM, KENNETH P.; and DAVID R. ANDERSON. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer-Verlag.

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.

COOPER, WILLIAM E.; and JOHN R. ROSS. 1975. World Order. *Papers from the Parasession on Functionalism*, ed. by Robin E. Grossman, James L. San, and Timothy J. Vance, 63–111. Chicago, IL: Chicago Linguistic Society.

DONOHUE, MARK. 2007. Word order in Austronesian from north to south and west to east. *Linguistic Typology* 11.349–391.

EHRET, KATHERINA LUISA; CHRISTOPH WOLK; and BENEDIKT SZMRECSÁNYI. 2014. Quirky quadratures: on rhythm and weight as constraints on genitive variation in an unconventional dataset. *English Language and Linguistics* 18.263–303.

EMBICK, DAVID. 2007. Linearization and Local Dislocation: Derivational Mechanics and Interactions. *Linguistic Analysis* 33.303–336.

EMBICK, DAVID; and ROLF NOYER. 2001. Movement Operations after Syntax. *Linguistic Inquiry* 32.555–595.

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

FERREIRA, VICTOR S.; and L. ROBERT SLEVC. 2007. Grammatical encoding. *The Oxford Handbook of Psycholinguistics*, ed. by Mark Gareth Gaskell, 453–469. Oxford, UK: Oxford University Press.

FOX, JOHN. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8.1–27.

FRANK, AUSTIN; and T. FLORIAN JAEGER. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Cogsci*. Washington, DC: CogSci.

FRENCH, KOLEEN MATSUDA. 1988. *Insights into Tagalog: Reduplication, infixation, and stress from nonlinear phonology*. Publications in Linguistics 84. Dallas, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.

FRENCH, KOLEEN MATSUDA. 1991. Secondary stress in Tagalog. *Oceanic Linguistics* 30.157–178.

FRISCH, STEFAN A.; JANET B. PIERREHUMBERT; and BROE, MICHAEL B. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228.

GARRETT, MERRILL F. 2000. Remarks on the architecture of language processing systems. *Language and the Brain: Representation and Processing*, ed. by Yosef Grodzinsky, Lewis P. Shapiro, and David Swinney, 3–69. Foundations in Neuropsychology. San Diego, CA: Academic Press.

GELMAN, ANDREW. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27.2865–2873.

GIBSON, EDWARD. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, ed. by Y. Miyashita, A. Marantz, and W. O'Neil, 95–126. Cambridge, MA: MIT Press.

GOLDSMITH, JOHN. 1976. Autosegmental phonology. Cambridge, MA: Massachusetts Institute of Technology ph.d. dissertation.

GRAFMILLER, JASON. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18.471–496.

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* 38.313–338.

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.

HANSSON, GUNNAR ÓLAFUR. 2001. Theoretical and Typological Issues in Consonant Harmony. University of California, Berkeley ph.d. dissertation.

HAWKINS, JOHN A. 1994. *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

HAYES, BRUCE. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: The University of Chicago Press.

HAYES, BRUCE; and TANYA STIVERS. 1996. *The phonetics of post-nasal voicing.*

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 and Linguistics* 11.437–474.

INKELAS, SHARON. 1990. *Prosodic Constituency in the Lexicon*. Outstanding Dissertations in Linguistics. New York, NY: Garland Publishing, Inc.

INTERNATIONAL PHONETIC ASSOCIATION. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. 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.

LEVELT, WILLEM. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

MARTIN, ANDREW T. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language* 87.751–770.

MCCARTHY, JOHN J. 1979. Formal problems in Semitic phonology and morphology. MIT.

MCCARTHY, JOHN J; and ALAN PRINCE. 1993. Generalized Alignment. *Yearbook of Morphology*, ed. by Geert E Booij and Jaap van Marle, 79–153. Dordrecht: Kluwer.

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.

MENN, LISE; and BRIAN MACWHINNEY. 1984. The Repeated Morph Constraint: Toward an Explanation. *Language* 60.519–541.

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 and Linguistics* 16.81–103.

MORTENSEN, DAVID. 2006. Logical and substantive scales in phonology. Berkeley, CA: University of California, Berkeley ph.d. dissertation.

ONISHI, KRISTINE H.; GREGORY L. MURPHY; and KATHRYN BOCK. 2008. Prototypicality in Sentence Production. *Cognitive Psychology* 56.103–141. doi:10.1016/j.cogpsych.2007.04.001.

PAK, MARJORI. 2008. The postsyntactic derivation and its phonological reflexes. Philadelphia, Pennsylvania: University of Pennsylvania ph.d. dissertation.

PATER, JOE. 1996. *NC. *Proceedings of the North East Linguistics Society 26*, ed. by Jill Beckman, 26:227–239. Amherst, Mass.: GLSA Publications.

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, 159–182. Cambridge: Cambridge University Press.

PRINCE, ALAN; and PAUL SMOLENSKY. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. RuCCS Technical Report. Piscateway, NJ: Rutgers University Center for Cognitive Science: Rutgers University.

QUIRK, RANDOLPH; SIDNEY GREENBAUM; GEOFFREY LEECH; and JAN SVARTVIK. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

RAFFELSIEFEN, RENATE. 1999. Phonological constraints on English word formation. *Yearbook of Morphology 1998*, ed. by Geert E Booij and Jaap van Marle, 225–287. Yearbook of Morphology 8. Springer.

ROSE, SHARON; and RACHEL WALKER. 2004. A Typology of Consonant Agreement by Correspondence. *Language* 80.475–531.

ROSENBACH, ANETTE. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81.613–644.

RYAN, KEVIN M. 2011. Gradient Syllable Weight and Weight Universals in Quantitative Metrics. *Phonology* 28.413–454.

RYAN, KEVIN M. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90.309–341.

SARKAR, DEEPAYAN. 2008. *"Lattice: multivariate data visualization with R"*. New York, NY: Springer.

SCHACHTER, PAUL; and FE T. OTANES. 1972. *Tagalog Reference Grammar*. Berkeley and Los Angeles, CA: University of California Press. http://www.ucpress.edu/op.php?isbn=9780520049437.

SCHLÜTER, JULIA. 2005. *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English*. Berlin, Germany: Mouton de Gruyter.

SEASITE. 2001. SEAsite Tagalog Online Dictionary (Large Tagalog Dictionary version). 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: The MIT Press.

SHIH, STEPHANIE; and JASON GRAFMILLER. 2011. Weighing in on End Weight. *The 84th Annual Meeting of the Linguistic Society of America*. Pittsburgh, PA.

SHIH, STEPHANIE S. 2014. Towards Optimal Rhythm. Stanford, CA: Stanford University phd dissertation.

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, 207–234. Berlin, Germany: De Gruyter Mouton.

SZMRECSÁNYI, BENEDIKT. 2004. On Operationalizing Syntactic Complexity. *7es Journ´ees internationales d'Analyse statistique des Donn´ees Textuelles*, 1031–1038. http://www.benszm.net/omnibuslit/Szmrecsanyi2004.pdf.

TEMPERLEY, DAVID. 2006. Minimization of dependency length in written English. *Cognition* 105.300–333.

TEMPERLEY, DAVID. 2009. Distributional Stress Regularity: A Corpus Study. *Journal of Psycholinguistic Research* 38.75–92.

TOPINTZI, NINA. 2006. Moraic Onsets. London, UK: University College London ph.d. 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, Gainesville, Florida, February 2000*, 191–218. Current Issues in Linguistic Theory. Amsterdam and Philadelphia: Benjamins.

TUCKER, MATTHEW A. 2011. The Morphosyntax of the Arabic Verb: Toward a Unified Syntax-Prosody. *Morphology at Santa Cruz: Papers in Honor of Jorge Hankamer*, 177–211. Santa Cruz, CA: Linguistics Research Center, University of California, Santa Cruz.

VIGLIOCCO, GABRIELLA; and ROBERT J. HARTSUIKER. 2002. The Interplay of Meaning, Sound, and Syntax in Sentence Production. *Psychological Bulletin* 128.442–472.

VOGEL, IRENE; and ISTVAN KENESEI. 1990. Syntax and semantics in phonology. *The Phonology-Syntax Connection*, ed. by Sharon Inkelas and Draga Zec, 339–364. Chicago: University of Chicago Press.

WASOW, TOM. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.

WOLF, MATTHEW. 2008. Optimal Interleaving: serial phonology-morphology interaction in a constraint-based model. University of Massachusetts Amherst.

WRIGHT, SAUNDRA K.; JENNIFER HAY; and TESSA BENT. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43.531–561.

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, 365–378. Chicago: University of Chicago Press.

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 & Linguistic Theory* 28.417–472.

ZWICKY, ARNOLD M. 1987. Suppressing the Z's. *Journal of Linguistics* 23.133–148.

ZWICKY, ARNOLD M.; and GEOFFREY PULLUM. 1986a. The principle of phonology-free syntax: introductory remarks. *Interfaces*, 32:63–91. Ohio State University Working Papers in Linguistics.

ZWICKY, ARNOLD M.; and GEOFFREY PULLUM. 1986b. Two spurious counterexamples to the Principle of Phonology-Free Syntax. *Interfaces*, 32:92–99. Ohio State University Working Papers in Linguistics.

Appendix A. QUANTIFIER ADJECTIVES

Limiters (string in corpus)

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

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[13]	'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'

---

[13] *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.'

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'