

VARIABLE AFFIX ORDER: GRAMMAR AND LEARNING

KEVIN M. RYAN

University of California, Los Angeles

While affix ordering often reflects general syntactic or semantic principles, it can also be arbitrary or variable. This article develops a theory of morpheme ordering based on local morphotactic restrictions encoded as weighted bigram constraints. I examine the formal properties of morphotactic systems, including arbitrariness, nontransitivity, context-sensitivity, analogy, and variation. Several variable systems are surveyed before turning to a detailed corpus study of a variable affix in Tagalog. Bigram morphotactics is shown to cover Tagalog and the typology, while other formalisms, such as alignment, precedence, and position classes, undergenerate. Moreover, learning simulations reveal that affix ordering under bigram morphotactics is subject to analogical pressures, providing a learning-theoretic motivation for the specific patterns of variation observed in Tagalog. I raise a different set of objections to rule-based approaches invoking affix movement. Finally, I demonstrate that bigram morphotactics is restrictive, being unable to generate unattested scenarios such as nonlocal contingency in ordering.*

Keywords: affix order, morphotactics, morphology, variation, Tagalog

Though the order in which morphemes are realized within a word is sometimes predictable on general syntactic or semantic grounds, ordering can also be subject to language-specific constraints, or exhibit free variation. Without denying the roles of the interfaces in affix ordering, this article focuses on the grammar and learnability of its arbitrary aspects. The theory of morphotactics must cover phenomena such as counterscopal or otherwise arbitrary ordering restrictions, nontransitivity (e.g. X-Y [**Y-X*] and Y-Z [**Z-Y*], but Z-X [**X-Z*]), context-sensitivity (e.g. X-Y-A [**Y-X-A*] but Y-X-B [**X-Y-B*]), gradient variation, learnability, and analogical effects (§6). For some of these phenomena, only a subset of logically possible patterns is found in human languages (e.g. context-sensitivity in ordering is always local). The theory should therefore also be restrictive, generating only possible patterns. I argue that morphotactics are encoded locally, as adjacency bigrams, while other proposals—such as morpheme alignment, precedence constraints, position classes, templates, and affix movement—critically undergenerate or overgenerate.

Gradient variation in affix ordering is a particularly stringent testing ground for morphotactic theories. I flesh out the typology of affix-order variation with cases from several languages before turning to the primary case study of this article, Tagalog's aspectual reduplicant (Schachter & Otones 1972 (hereafter S&O), Carrier 1979, Conдоравди & Kiparsky 1998, Rackowski 1999, Mercado 2007, Skinner 2008). Building on previous accounts with a corpus study, I discuss the productive characteristics of the system, including its gradience (not all options are equally good) and the sorts of morphological contingencies the morphotactic grammar must address. A sample of Tagalog data follows in 1–4 as a brief illustration of such gradience and contingency (see §3 for details). First, individuals freely place aspectual 'RED' before or after the causative prefix *pa*, as in 1. The corpus incidence of the affix matrix (with RED in any position) and percentages of the variants appear to the right. As discussed in §3, the variants' proportions are generally consistent across roots.

* I wish to thank Kie Zuraw, Bruce Hayes, Hilda Koopman, Laura McPherson, Joe Pater, Colin Wilson, audiences at UCLA, UC Berkeley, and the LSA annual meeting, the editors, and three anonymous referees for their insightful questions, criticisms, and suggestions. I also thank Larry Hyman and Sam Mchombo for their assistance with the Chichewa data. All faults are my own. This material is based on work supported by a National Science Foundation Graduate Research Fellowship.

- (1) pa-RED-ROOT-an (64.5%)
 ~ RED-pa-ROOT-an (35.5%) ($N = 212,140$)

When *pag* ‘transitive’ is added outside of *pa* as in 2, however, the position after *pa* is almost never employed. (Variants whose corpus proportions round to 0.0% are not shown.) Third, the position between *pag* and *pa* continues to be the preferred variant (though more weakly so) when two additional prefixes, *ma* ‘ability’ and *ʔi* ‘object topic’, are added to the prefix string in 2 to make 3. Finally, when *ʔi*, the second prefix in 3, is replaced with *ka* ‘telic’ in 4, a higher position in the verb, between *ma* and *ka*, is preferred.

- (2) pag-RED-pa-ROOT (100%) ($N = 880,685$)
 (3) ma-ʔi-pag-RED-pa-ROOT (96.0%)
 ~ ma-ʔi-RED-pag-pa-ROOT (4.0%) ($N = 3,073$)
 (4) ma-ka-pag-RED-pa-ROOT (25.4%)
 ~ ma-RED-ka-pag-pa-ROOT (74.6%) ($N = 43,623$)

I demonstrate that these sorts of ‘free’ variation in Tagalog and other languages can be modeled by a morphotactic grammar containing weighted bigram constraints, which effectively encode the relative coherences of pairs of cooccurring morphemes. Output candidates exhibiting all ordering permutations are evaluated by the constraint set for their global morphotactic harmonies, which can be translated into probabilities. In a learning simulation based on the comprehensive corpus data, the bigram learner closely matches both the categorical and gradient properties of its Tagalog training data.

One important finding of this article is that the set of marked (grammatical but not preferred) variants in Tagalog is predictable from the set of unmarked (preferred) variants. When the bigram learner is trained on the Tagalog corpus with all marked variants removed, the learner overgenerates variation in patterns that resemble the unseen actual corpus, both in terms of the positions that are overgenerated and the degrees to which they are overgenerated. I am therefore able to motivate the evolution and stability of such a complex morphotactic system. The variation, I suggest, is driven by the overgeneralization of local morphotactics, a kind of analogy in affix ordering that I refer to as morphotactic extension. In sum, bigram morphotactics is supported not only by its descriptive adequacy, covering the typology where other theories fall short, but also by its predictive power in modeling analogical effects during learning.

The article is organized as follows. I first focus primarily on description, establishing the typology of affix ordering systems and examining a corpus study of a variable affix in Tagalog. The theory of morphotactic constraints is then developed, including learning simulations and discussion of morphotactic extension as an explanation of free variation in ordering. Following this, I argue against other theories of affix ordering, including both rule-based accounts and other constraint formalisms. Finally, I discuss nontransitivity, and then address the restrictiveness of bigram morphotactics.

1. FOUR TYPES OF AFFIX ORDERING. Consider a verb with two derivational affixes. For simplicity, let us assume that the two affixes surface adjacently on the same side of the root; more complex situations are treated in the following sections. Four ordering scenarios are possible. First, the ordering of the affixes might be determined by semantic scope. For example, both Quechua verbs in 5 contain causative *chi* and reciprocal *na* (Muysken 1986).¹ In 5a the reciprocal scopes over the causative (i.e. the causation is re-

¹ I employ the following morpheme gloss abbreviations in this article: 1 (etc.): first (etc.) person, ABIL: ability, AGR: subject agreement, APPL: applicative, ASP: aspect (particularly, Tagalog’s contemplated aspect),

ciprocal). In 5b the causative scopes over the reciprocal (i.e. the seeing is reciprocal). In both cases, the higher-scoping morpheme is more offset from the root, as predicted by ordering principles based on semantic scope (e.g. Muysken 1981a,b, Bybee 1985, Psetsky 1985, Rice 1993, 2000) or on affix ordering reflecting the order of syntactic operations (Baker 1985, cf. Alsina 1999).

- (5) a. riku-chi-na-n-ku
 see-CAUS-REC-3-PL
 ‘they make each other see (something)’
 b. riku-na-chi-n-ku
 see-REC-CAUS-3-PL
 ‘they make (them) see each other’

Second, the order might be fixed, even if the scope is reversible. In Luganda, for instance, a passivized causative and a causativized passive can both only be realized with causative *es* preceding passive *ebw* (e.g. the stem *nyw-es-ebw* ‘drink-CAUS-PASS’ is used for both ‘[be [made to drink]]’ and ‘[make [be drunk]]’; McPherson & Paster 2009). Other fixed orderings (despite reversible scope) have been discussed in Bantu and elsewhere (e.g. Hyman 1994, 2003, Good 2003, 2005, 2007, McFarland 2005, Paster 2006a,b, Caballero 2008, 2011). A second example, from Mapuche (Smeets 1989:348), is given in 6 in which the simulative and negative are fixed in that order regardless of whether the negative scopes over or under the simulative.

- (6) a. pe-w-faluw-la-e-y-u
 see-REFL-SIM-NEG-IND.OBJ-IND-AGR
 ‘I did not pretend to see you (sg.)’
 b. pe-w-faluw-la-e-y-u
 see-REFL-SIM-NEG-IND.OBJ-IND-AGR
 ‘I pretended not to see you (sg.)’

Third, the affixes might be freely ordered for one scope but fixed (according to scope) for the other. Hyman (2003:250) refers to this situation as ‘asymmetric compositionality’ and claims it is exhibited by the Chichewa causative *its* and reciprocal *an* suffixes. For a causativized reciprocal semantics (7a), either order is grammatical (Hyman 2003; full examples in 7 and 10 from Larry Hyman and Sam Mchombo, p.c.). For a reciprocalized causative (7b), only the scopal order CAUS-REC is available. For similar cases in Choguita Rarámuri, see Caballero 2011.

- (7) a. a-ku-máng-íts-a:n-a ~ a-ku-máng-án-i:ts-a
 3PL-PROG-tie-CAUS-REC-FV 3PL-PROG-tie-REC-CAUS-FV
 ‘cause to tie each other’
 b. a-ku-máng-íts-a:n-a, *a-ku-máng-án-i:ts-a
 3PL-PROG-tie-CAUS-REC-FV 3PL-PROG-tie-REC-CAUS-FV
 ‘cause each other to tie’

Fourth, the two affixes might be freely ordered for both scopes. For example, Fuuta Tooro Pulaar causative *in* and applicative *ir* can occur in either order for both causativized applicatives and applicativized causatives, as in 8 (Paster 2006b:182). Many other cases of optionality in affix ordering have been discussed in several lan-

CAUS: causative, DET: determiner, FV: Bantu final vowel suffix, IND: indicative, IND.OBJ: indirect object, LT: locative topic, NEG: negative, NS: nonsingular, OT: object topic, PASS: passive, PL: plural, PROG: progressive, REC: reciprocal, RED: reduplicant (particularly, Tagalog’s aspectual reduplicant), REFL: reflexive, SG: singular, SIM: simulative, SOC: social, TEL: telic, TH: thither (Smeets 1989), TRANS: transitive.

guage families (e.g. Albó 1964:appendix 5, Stevens 1971, Herrero & Sánchez de Lozada 1978:227ff., Sabimana 1986, Muysken 1988, Smeets 1989:360, van de Kerke 1996:§3, Luutonen 1997, Good & Yu 1998, 2005, Rackowski 1999, Blevins 2001:118, Hyman 2003, Watters 2006:62, Beck 2007, Bickel et al. 2007, Joseph 2007, McFarland 2007, Buell et al. 2008, Caballero 2008, 2011).

- (8) a. o irt-ir-in-ii ~ irt-in-ir-ii kam supu o kuddu
 3SG stir-APPL-CAUS-PAST stir-CAUS-APPL-PAST 1SG soup DET spoon
 ‘he made me stir the soup with a spoon’ (I used a spoon)
- b. o irt-ir-in-ii ~ irt-in-ir-ii kam supu o lafi
 3SG stir-APPL-CAUS-PAST stir-CAUS-APPL-PAST 1SG soup DET knife
 ‘he made me stir the soup with a knife’ (he used a knife)

This fourfold typology is summarized in Table 1, which shows the possible output(s) for both possible semantic scopes in each type of grammar. The grammars are modeled by rankings of three constraints, namely, two morphotactic constraints (e.g. ‘X-Y’ prefers an output with X-Y; for more details, see §4) and a SCOPE constraint pressuring for scopal ordering (Condoravdi & Kiparsky 1998, Paster 2006a,b, Aronoff & Xu 2010; similar violable constraint proposals include MIRROR in Hyman 2003, Wolf 2008, and McCarthy 2011; LINEAR CORRESPONDENCE in Ackema & Neeleman 2004, 2005; and LINEARITY in Horwood 2002). Tilde indicates indeterminate or free ranking; ‘>>’ strict domination; ‘[[... X]Y]’ an input in which Y scopes over X. Assume that other morphotactics guarantee that the root is always initial, so that X and Y are always suffixes.

	SCENARIO	RANKING	OUTPUT(S) FOR /[[[root]X]Y]/	OUTPUT(S) FOR /[[[root]Y]X]/
i.	compositional	SCOPE >> X-Y ~ Y-X	root-X-Y	root-Y-X
ii.	fixed (a)	X-Y >> SCOPE ~ Y-X	root-X-Y	root-X-Y
	fixed (b)	Y-X >> SCOPE ~ X-Y	root-Y-X	root-Y-X
iii.	asymmetric (a)	X-Y ~ SCOPE >> Y-X	root-X-Y	root-X-Y, root-Y-X
	asymmetric (b)	Y-X ~ SCOPE >> X-Y	root-X-Y, root-Y-X	root-Y-X
iv.	free	X-Y ~ Y-X >> SCOPE	root-X-Y, root-Y-X	root-X-Y, root-Y-X

TABLE 1. Basic ordering typology for adjacent affixes.

Finally, note that the ordering scenario associated with a pair of morphemes is generally unpredictable from the semantics of those morphemes. For instance, causative and reciprocal are scopally ordered in Quechua (5), asymmetrically compositional in Chichewa (7), and fixed in Luganda (McPherson & Paster 2009). There is even variation across Quechua dialects on this point: ‘The highland dialects ... allow reciprocal to both precede and follow causative; this does not appear to be possible in the jungle dialects’ (Muysken 1981a:470). Moreover, different pairs of affixes within a language might exhibit different ordering facts. In Fuuta Tooro Pulaar, for example, *in* and *ir* are freely ordered (8), *id* ‘comprehensive’ and *in* vary according to scope, and *it* ‘repetitive’ and *ir* are fixed in that order regardless of scope (Paster 2006b).

2. FREE VARIATION IN AFFIX ORDERING. In §1 I mentioned the case of two Pulaar affixes being freely ordered regardless of scope. In this section I describe four additional scenarios of free variation in affix order. This is only an introductory survey; analysis and empirical details of free variation, particularly in Tagalog, are pursued in the following sections.

First, a morpheme might vary freely between nonadjacent positions. Examples from five languages are given in 9. In each case, the boxed morpheme surfaces freely (i.e.

synonymously) before or after a two-morpheme sequence, but cannot interrupt it. The intervening bigrams are clearly compositional, bimorphemic strings in each language. Each of the morphemes can occur with the glossed semantics independently of its neighbor (sources cited in 9). Furthermore, in each case, the source explicitly rejects any other ordering options for the given meaning. Tagalog RED is an aspectual reduplicant, to be treated in detail in the following sections, where I argue that its placement is morphologically determined (§3) and suggest a motivation for this pattern (§6). On the semantics of Tagalog *ka* ‘telic’ and *pag* ‘transitive’, see Travis 1996, 2007, and Rackowski 1999.

(9) a. Chichewa

a-ku-máng-il-an-i:ts-a
 3PL-PROG-tie-APPL-REC-CAUS-FV
 ~ a-ku-máng-its-il-a:n-a
 ‘they make each other tie with’ (a rope) (Hyman 2003:273)

b. Mapuche

llepipu-ñma-fal-ye-nge-me-y
 request-IND.OBJ-FORCE-PL-PASS-TH-AGR
 ~ llepipu-ye-ñma-fal-nge-me-y
 ‘they have to be requested many things’ (Smeets 1989:361)

c. Tagalog

ma-RED-ka-pag-pa-bili
 ABIL-ASP-TEL-TRANS-CAUS-buy
 ~ ma-ka-pag-red-pa-bili
 ‘will be able to make buy’ (S&O:362)

d. Dakar Wolof

lekk-oon-na-a
 eat-PAST-FINITE-1SG
 ~ lekk-na-a-woon
 ‘I was not eating’ (Buell et al. 2008:12)

e. Chumbivilcas Quechua

kiki-la-n-kuna
 self-just-3-PL
 ~ kiki-n-kuna-la
 ‘just themselves’ (Muysken 1981b:295)

Second, an affix might vary freely among three or more positions within a span of morphemes that is otherwise fixed in order. Tagalog’s aspectual reduplicant exemplifies this case (10). In a verb of the form *ma-ʔi-pa-ROOT* ‘ABIL-OT-CAUS-ROOT’, RED can intervene between *ma* and *ʔi*, *ʔi* and *pa*, or *pa* and the root, but *ma*, *ʔi*, *pa*, and the root are fixed in that order (S&O:362). Tagalog researchers are unanimous that the optionality in RED placement here and elsewhere is free of any semantic effects. Rackowski, for one, concludes that ‘native speakers of Tagalog produce all the different variants in their own speech, and these do not perform differently on any aspectual tests that would discriminate between the different readings’ (1999:4). Good and Yu (1998) describe another possible case of this type of variation in Turkish.

- (10) ma-RED-ʔi-pa-ʔabot ~ ma-ʔi-RED-pa-ʔabot ~ ma-ʔi-pa-RED-ʔabot
 ABIL-ASP-OT-CAUS-talk ABIL-OT-ASP-CAUS-talk ABIL-OT-CAUS-ASP-talk
 ‘will talk to someone’

Third, two affixes might be freely swappable (as in Pulaar) but only in certain morphological contexts. For example, Tagalog RED and *ka* ‘telic’ can be freely exchanged

before the root (11a) but not before *pag* ‘transitive’ (11b) (S&O; below, §3). Chichewa also exemplifies this case. As seen in 7 above, REC and CAUS can surface in either order for a causativized reciprocal ([[root-R]C]) with no other derivational morphology. For a causativized reciprocalized applicative ([[root-A]R]C]), however, the scopal order APPL-REC-CAUS is licit but *APPL-CAUS-REC is not. Thus, REC and CAUS are freely swappable except following APPL (data from Hyman 2003:273).

- (11) a. ma-RED-ka-tulong ~ ma-ka-RED-tulong
 ABIL-ASP-TEL-help ABIL-TEL-ASP-help
 ‘will be able to help’
 b. ma-RED-ka-pag-trabaho, *ma-ka-RED-pag-trabaho
 ABIL-ASP-TEL-TRANS-work ABIL-TEL-ASP-TRANS-work
 ‘will be able to work’

Fourth, a group of more than two adjacent affixes might all be freely ordered with respect to each other. Bickel and colleagues (2007:44) describe this situation in Chintang. For example, the three prefixes in 12, namely, *ma* ‘negative’, *u* ‘3rd-person nonsingular agent’, and *kha* ‘1st-person nonsingular primary object’, are grammatical in all six permutations with no meaning differences. This type of variation is also exhibited by certain sets of suffixes in (e.g. Cochabamba) Quechua (van de Kerke 1996:§3).

- (12) u-kha-ma-cop-yokt-e
 3NS.AGENT-1NS.P-NEG-see-NEG-PAST
 ~ u-ma-kha-cop-yokt-e
 ~ kha-u-ma-cop-yokt-e
 ~ ma-u-kha-cop-yokt-e
 ~ kha-ma-u-cop-yokt-e
 ~ ma-kha-u-cop-yokt-e
 ‘they didn’t see us’

3. THE TAGALOG CONTEMPLATED ASPECT MORPHEME. The contemplated (unrealized) aspect of Tagalog is marked by a long, weakly stressed CV: reduplicant prefix. I refer to this morpheme as RED (corresponding to RED_a in S&O), though it should be understood throughout that I am only treating one of a number of reduplicative morphemes in Tagalog, which do not all behave in the same way. If a verb has no other prefixes, RED immediately precedes the root, for example, *sà:-salita?-ín* ‘RED-talk-OT’ (OT being ‘object topic’; topic markers indicate the role of the topic-marked argument in the sentence).

The phonological facts support the analysis of RED as a prefix as opposed to an infix (i.e. *sà:-sa-lita?-ín*). First, cluster simplification, if asymmetric, obtains in the first copy, not the second, for example, *trabahuh-in* ‘work-LT’ yields *tà:-trabahuh-in* with RED. Second, segmental nativization, if asymmetric, obtains in the first copy, not the second; for example, *mag-θænkyu* ‘thank’ can yield *mag-tè:θænkyu* but not **mag-θè:-tænkyu* (Zuraw 1996:8; see McCarthy & Prince 1995 on base vs. reduplicant markedness, cf. Inkelas & Zoll 2005). Finally, stress and length are neutralized in the first copy, since RED is always long and stressed, while remaining faithful in the second.

In verbs with one or more prefixes, RED often varies freely in position for individual speakers. For example, in the verb *pa-buks-án* ‘CAUS-open-LT’ (LT being ‘locative topic’) in 13, RED can either precede or follow *pa* ‘CAUS’. Either way, the aspect scopes over the causative. The options for RED placement are always synonymous (see §2). (All judgments come from S&O and a corpus study, described below.)

- (13) a. [pà:]pa-buks-án ‘RED-CAUS-open-LT’
 b. pa-[bù:]buks-án ‘CAUS-RED-open-LT’
 ‘will cause to open’

In some verbs with three or more prefixes, RED can take three or more positions, as in 10 above in §2. But other verbs with multiple prefixes are more restrictive about where RED can be situated, for example, *ma-ki-pag-pa-sayá* ‘ABIL-SOC-TRANS-CAUS-be.happy’ (SOC being ‘social’, TRANS ‘transitive’), in which RED is only used before *ki*. It is ungrammatical or degraded in all other positions. But when *ki* is replaced by *ka* ‘telic’, as in 14, the position between *pag* and *pa*, unavailable in *ma-ki-pag-pa-sayá*, becomes a grammatical option.

- (14) a. *mà:-ma-ka-pag-pa-sayá ‘RED-ABIL-TEL-TRANS-CAUS-be.happy’
 b. ma-kà:-ka-pag-pa-sayá ‘ABIL-RED-TEL-TRANS-CAUS-be.happy’
 c. *ma-ka-pà:-pag-pa-sayá ‘ABIL-TEL-RED-TRANS-CAUS-be.happy’
 d. ma-ka-pag-pà:-pa-sayá ‘ABIL-TEL-TRANS-RED-CAUS-be.happy’
 e. ?ma-ka-pag-pa-sà:-sayá ‘ABIL-TEL-TRANS-CAUS-RED-be.happy’
 ‘will be able to make happy’

Although I have exemplified only particular roots in the Tagalog data thus far, the facts about the placement of RED are essentially the same regardless of which root is used with the given prefixes. For one, this is how the descriptive grammarians state the facts. According to the rules of S&O (p. 362) and the paradigms of Ramos & Bautista 1986, RED is grammatical immediately before *ka* or *pa* and illicit elsewhere in 14. Samuels (2006) likewise rejects 14a and 14c but implies that 14e is licit. In my corpus data (below), the order in 14e is encountered 0.04% of the time for this prefix string.

Corpus data support the judgments of the descriptive grammarians. The bar chart in Figure 1 shows the Google hits (www.google.com, retrieved May 2009) for the prefix string *ma-ka-pag-pa* with RED in each of the five positions listed in 14. Each bar represents the hits for the position indicated summed over 321 different Tagalog roots queried with this prefix string. For example, the ‘RED-ka’ bar in Fig. 1 represents the sum of the results for *makakapagpabili* (188 hits), *makakapagpakita* (1,310 hits), and so forth for the remaining 319 roots in the pilot lexicon, making for 32,488 total hits. Searches were automated by a Perl script with the LWP::UserAgent module (distributed by www.cpan.org). This chart reveals that for *ma-ka-pag-pa*, the position before *ka* is preferred almost three to one over the position before *pa*, while all other positions, including root reduplication, are rarely if ever encountered in the corpus.

FIGURE 1. RED placement in the base *ma-ka-pag-pa*-ROOT.

Furthermore, the aggregate preference for RED-*ka* depicted in Fig. 1 is relatively stable across roots ($r = 0.84$ for the 172 roots of the 321 queried that had nonzero results

for *ma-ka-pag-pa* plus RED). This rough consistency across roots is illustrated by the jittered one-dimensional scatterplot in Figure 2. Each circle represents one of the 172 roots attested with *ma-ka-pag-pa* plus RED in any position. For example, the largest circle represents *saya* ‘be happy’, for which there were 13,495 tokens with *ma-ka-pag-pa* plus RED. The x-coordinate of the circle center represents the proportion of the time that RED is placed before *ka* when that root is prefixed with *ma-ka-pag-pa* and RED; for *saya* it is 0.748. The size of each circle is proportional to the sample size on which that proportion is based. Thus, x-coordinates of larger circles are expected to be closer to their underlying propensities than those of smaller circles, which reflect smaller and therefore less reliable samples. The weighted mean of the circles in Fig. 2 is 0.745, the same as that given in Fig. 1.

FIGURE 2. Average pre-*ka* placement for 172 roots.

Table 2 shows the Web-based corpus results for RED in each prefixal position in twenty-nine prefix strings, organized by the number of prefixes. In this table, ‘pos. 0’ is always the position before the first prefix, ‘pos. 1’ is after the first prefix, and so forth. Incidences for each position are given as percentages rounded to the nearest tenth. For example, the first row indicates that in verbs of the form *ka*-ROOT-*an*, RED precedes *ka* 31.9% of the time and follows it 68.1% of the time. ‘Token *N*’ is the total incidence of the affix matrix plus RED in the corpus. ‘Type *N*’ is the number of roots attested in the corpus with the given affix matrix plus RED in any position. In total, 321 roots were queried with RED in each position in each prefix string. In some rows the type count exceeds 321 because these rows combine counts for affix matrices with different suffixes. For instance, the ‘*pa*- + {*-an*, *-in*}’ row has a type *N* of 337 because it sums the results for ‘*pa*- + *-an*’ (161 roots) and ‘*pa*- + *-in*’ (176 roots). As this practice implies, the choice of suffix has no effect on RED placement in the prefix string. In this case, for instance, there is no significant difference between the incidence of RED before *pa*- in the *-an* types vs. the *-in* types, Welch’s $t(615) = -0.58$, $p = 0.56$. Finally, to increase the scope and reliability of the study, all *m*-initial forms were also queried with nasal-substituted *n*-initial variants (e.g. *ma-ka-* includes results for *na-ka-*). Nasal substitution (see Kroeger 1993 on semantics, Zuraw 2010 on phonology) and RED together indicate imperfective aspect.²

Percentages in Table 2 are based on token frequencies. For example, *ma-ka-pag-pa*-ROOT has a token *N* of 43,623 (summing over all 172 roots found with this prefix string plus RED), so the 74.6% for ‘pos. 1’ corresponds to approximately 32,500 tokens. It would also be possible to obtain these percentages by averaging across types (i.e.

² Morpheme glosses are as follows: *-an* ‘locative topic’, *-in* ‘object topic’, *ka-* ‘telic’, *ki-* ‘social’, *ma-* ‘ability’, *mag-* ‘actor topic’, *pa-* ‘causative’, *pag-* ‘transitive’, *si-* ‘plural’, *ʔi-* ‘object topic’.

PREFIXES	AFFIX MATRIX	TOKEN <i>N</i>	TYPE <i>N</i>	POS. 0	POS. 1	POS. 2	POS. 3	POS. 4
1	ka- + -an	165,043	156	31.9%	68.1			
	ma- + -an	934,762	204	0	100			
	pa- + {-an, -in}	510,090	337	35.5	64.5			
	pag- + {-an, -in}	527,618	339	0.6	99.4			
2	?i-ka-	274,986	175	0	28.5	71.5		
	?i-pa-	951,363	262	0	83.9	16.1		
	?i-pag-	306,153	180	0	1.5	98.5		
	ka-pag-	15,612	79	37.4	0	62.6		
	mag-ka-	559,113	176	0	99.9	0.1		
	mag-pa-	970,285	275	0	100	0		
	mag-si-	14,429	114	0	99.9	0.1		
	ma-?i-	653,862	257	0	13.5	86.5		
	ma-ka- + { \emptyset , -an, -in}	6,053,626	360	0	97.2	2.8		
	ma-ki-	275,096	237	0	95.5	4.5		
	ma-pa- + { \emptyset , -an, -in}	706,361	318	0	96.0	4		
	ma-pag- + { \emptyset , -an, -in}	73,100	223	0	5.7	94.3		
	pag-pa- + { \emptyset , -an, -in}	883,858	328	0	100	0		
	pa-ki-	125,622	121	0.6	92.2	7.2		
3	?i-ka-pag-	386	26	0	2.8	0.3	96.9	
	?i-pag-pa-	51,776	46	0	6.8	93.2	0	
	?i-pa-ki-	762	17	0	17.9	62.6	19.5	
	mag-si-pag-	2,724	74	0	100	0	0	
	ma-?i-pa-	66,418	121	0	0.3	90.9	8.8	
	ma-?i-pag-	17,152	61	0	0.1	1.5	98.4	
	ma-ka-pag-	203,181	231	0	61.8	0	38.2	
	ma-ki-pag-	28,492	54	0	99.9	0	0.1	
4	?i-ka-pag-pa-	133	8	0	1.5	0.8	97.7	0
	ma-?i-pag-pa-	3,073	15	0	0	4.0	96.0	0
	ma-ka-pag-pa-	43,623	172	0	74.6	0	25.4	0

TABLE 2. RED placement in twenty-nine prefix strings.

roots), but the results would be essentially the same. For example, when the root with *ma-ka-pag-pa* is *tawa*, the pre-*ka* position is used 71.7% of the time. For *tayo*, it is 78.6% of the time. Averaging over the 172 percentages for roots attested in this frame, RED is found in the pre-*ka* position 73.5% of the time, a small difference from the 74.6% figure based on the overall token count for that position. Results would be similar for all percentages given ($r = 0.964$ for type- vs. token-derived percentages).

Most sources agree that the facts about RED's placement cannot be determined by phonology (S&O, Carrier 1979, Condoravdi & Kiparsky 1998, Rackowski 1999, etc.). First, unlike Tagalog's infixes, RED never interrupts a morpheme. Second, the preferences in placing RED are apparently insensitive to the prosodic properties, for example, stress pattern, of the root and its affix matrix. For example, nearly phonologically identical pairs of prefix strings can exhibit markedly different patterns in RED placement, for example, *ma-ki-pag*, in which RED almost never follows *pag*, contrasting with *ma-ka-pag*, in which RED often follows *pag*. Third, the rules need to be stated over morphemes, as in S&O, rather than over segments or prosodic positions. For instance, RED cannot precede word-initial *ma* 'ability'. This cannot be because RED is avoided word-initially; for example, it can precede (otherwise) initial *pa*. Nor is there a dispreference for the reduplication of *m* or *ma* as opposed to other strings. RED can felicitously precede *ma* if it is part of the root. More complex morphological dependencies are treated in §4.

To briefly address one particular phonological proposal, Inkelas (2000) and Inkelas and Zoll (2005) (see also French 1988, Booij & Lieber 1993, Cole 1994, Downing 1998) suggest that RED variation in Tagalog can be explained as optionality between root and second-syllable reduplication. Although they consider a variety of reduplicants, some of their examples involve the aspectual reduplicant, implying that the analysis is meant to cover it. A prefix string such as *ti-pag-pa* demonstrates the untenability of this proposal for aspectual RED. In this string, the position between *pag* and *pa* is preferred for RED, accounting for 93% of some 50,000 corpus instances. This is neither the second syllable nor preroot position. Nor is it possible to interpret *pa* as part of the (prosodic) root, because in other prefix strings, RED can intervene between *pa* and the root.

4. MORPHOTACTIC CONSTRAINTS. I argue that the sorts of morphological variation described in §§2 and 3 are the result of tensions between different morphotactic constraints competing in the grammar. I propose that these constraints are local in the sense that each evaluates only a pair of adjacent morphemes. In this section, I sketch how bigram morphotactics can implement the four variable scenarios in §2. I argue against other possible formalisms in §§7 and 8.

A bigram constraint X-Y, in which X and Y are (classes of) morphemes, can be taken to penalize each instance of X not immediately followed by Y (cf. local selectional restrictions, e.g. Fabb 1988).³ The ranking of these constraints motivates ordering restrictions, as in 15, in which X-Y-Z is the only grammatical output for an input comprising X, Y, and Z. This OPTIMALITY THEORY (OT; Prince & Smolensky 2004 [1993]) tableau includes all six ordered pairs of morphemes as constraints and all six orders of the three morphemes as candidates. Dashed vertical lines indicate indeterminate or freely variable rankings.

(15) Tableau illustrating ranked bigram constraints

	X-Y	Y-Z	X-Z	Y-X	Z-Y	Z-X
a. ☞ X-Y-Z			*	*	*	*
b. X-Z-Y	*!	*		*		*
c. Y-X-Z	*!	*			*	*
d. Y-Z-X	*!		*	*	*	
e. Z-X-Y		*!	*	*	*	
f. Z-Y-X	*!	*	*			*

It is also possible that the learner posits constraints only for bigrams that it has actually encountered, in which case only the first two constraints in 15 would be present.

³ For the purposes of this article, I interpret X-Y as assigning a violation iff a candidate contains X not immediately followed by Y. Other interpretations favoring candidates containing X-Y over those not containing X-Y would work just as well for the data treated here. I remain neutral as to whether these constraints are part of a combined phonological/morphological constraint-based component (along the lines of Wolf 2008 and McCarthy 2011) or part of an autonomous morphological component (cf. distributed morphology), among other possibilities. The present proposal is compatible with either architecture, given some flexibility about the precise interpretation of the constraints (e.g. whether they refer to abstract morphemic indices or to phonological material licensed by morphemes).

For simplicity, I omit candidates with morpheme duplication or deletion, as these can be ruled out by constraints that are unviolated in all the data considered in this article. On constraints requiring morpheme realization, see Noyer 1993, Samek-Lodovici 1993, Rose 1997, Walker 1998, Kurisu 2001, Ussishkin & Wedel 2002, MacBride 2004, Wolf 2008, and many others; on ruling out gratuitous duplication of morphemes, see Noyer 1993, Peterson 1994, Jaker 2006, Inkelas & Caballero 2008, Wolf 2008, among others. Phonological constraints such as MAX and DEP (McCarthy & Prince 1995) could also be used to ensure biunique mapping in the cases analyzed in this article. Furthermore, under bigram morphotactics, circumfixes can be treated as pairs of morphosyntactically or semantically interdependent affixes.

Candidates with morpheme duplication (etc.) need to be generated, however, since multiple exponence is known to occur (e.g. Inkelas & Caballero 2008, Caballero 2010, and numerous references therein). To give one relevant example, Pima compound pluralization (Munro & Riggle 2004, Riggle & Wilson 2004) superficially resembles Tagalog in that a reduplicant is free to occur in multiple morphological positions in the word, as in 16a–c (leaving RED abstract for clarity). Unlike Tagalog, however, RED can also be simultaneously realized in any combination of the positions 16d–g. One potential analysis of 16 is that RED duplication is motivated by local morphotactics: if bigram constraints such as RED-root₁, RED-root₂, and RED-root₃ are freely ranked with a constraint assigning a penalty for each copy of RED, all of the options in 16 are generated (this was confirmed using OT-Help, Becker et al. 2007; to rule out the candidate with no copy of RED realized, a constraint such as MAX(morph) ‘assign a penalty if a morpheme is unrealized’ can be undominated). This line of analysis would also permit giving the constraints different weights to account for gradient preferences among the options in 16. I leave exploring the typology of possible interactions between morphotactics and multiple exponence to future work.

- (16) Plurals of the Pima compound ¹us-kàlit-váinom ‘wagon knife (lit. tree-car-knife)’
- a. RED-¹us-kàlit-váinom
 - b. ~¹us-RED-kàlit-váinom
 - c. ~¹us-kàlit-RED-váinom
 - d. ~RED-¹us-RED-kàlit-váinom
 - e. ~RED-¹us-kàlit-RED-váinom
 - f. ~¹us-RED-kàlit-RED-váinom
 - g. ~RED-¹us-RED-kàlit-RED-váinom

I first address the case of a morpheme varying freely between nonadjacent positions. Recall, for example, that Tagalog RED can optionally align on either side of the *ka-pag* prefix bigram, but cannot interrupt it. The fact that the variable morpheme is a reduplicant here is apparently not synchronically crucial (regardless of the diachrony, cf. §6); after all, all of the scenarios involving RED discussed in §2 were also exemplified by fixed-melody morphemes in other languages. This case can be modeled by a highly ranked (or weighted) *ka-pag* bigram, representing the speaker’s knowledge of the relative coherence (unsplittability) of these two affixes. At the same time, the variation of RED can be motivated by two or more variably ranked (or closely weighted) bigrams pressuring for both the pre-*ka* (and/or post-*ma*) and pre-*pa* (and/or post-*pag*) positions, as in 17. To save space, ‘√’ is used for ‘ROOT’ in tableaux. The numbers are explained in the following paragraphs.

(17) Tableau illustrating variation between nonadjacent positions

			RED- ka	pag- RED	ka- pag	pag- pa	pa- RED	RED- pa	ma- RED	ma- ka	pa- √	RED- ka
GEN'D	SCORE	CANDIDATE	11.1	10.9	9.6	7.9	7.6	7.2	6.9	6.7	6.5	2.7
a.	74.51%	35.1	ma-RED-ka-pag-pa-√	*			*	*		*		*
b.	25.45%	36.2	ma-ka-pag-RED-pa-√	*		*	*		*			*
c.	0.04%	42.6	ma-ka-pag-pa-RED-√	*	*			*	*		*	
d.	0.00%	56.0	ma-ka-RED-pag-pa-√	*	*	*		*	*	*		*
e.	0.00%	63.7	ma-RED-pag-ka-pa-√	*	*	*	*	*		*		*

In HARMONIC GRAMMAR (e.g. Smolensky & Legendre 2006, Pater 2009, Potts et al. 2010), constraints are assigned real-number weights and each candidate’s score is the sum of its weighted constraint violations. I employ a type of harmonic grammar termed MAXIMUM ENTROPY (maxent) grammar (e.g. Johnson 2002, Goldwater & Johnson 2003, Wilson 2006, Hayes & Wilson 2008) to model gradient variable data. In tableau 17, the weight of each constraint is given under its label and the score of each candidate (lower being better) is given to its left. Scores are computed as in 18, in which i ranges over constraints, $C_i(x_n)$ gives the violations incurred by candidate x_n on constraint C_i , and w_i is C_i ’s weight.

$$(18) S(x_n) = \sum_{i=1}^M w_i C_i(x_n)$$

Weights were computed using maxent learning software by Wilson and George (2008), which draws on the CONJUGATE GRADIENT method of Press et al. 1992 (no smoothing term is employed here). The learner is given the actual Tagalog output frequencies, from which it weights the constraints so as to maximize the fit between generated and observed frequencies. The generated (‘gen’d’) percentages in the first column in 17 were computed as follows. First, $H(x_n)$, the ‘maxent value’ of a candidate x_n , is given by raising the constant e (approximately 2.718) to the negation of x_n ’s score. Then, to get $P(x_n)$, the probability of x_n , $H(x_n)$ is divided by the sum of the maxent values of all candidates, as in 20. Note that like Goldwater & Johnson 2003 and Wilson 2006 but unlike Hayes & Wilson 2008, maxent values are translated into probabilities on a per-input basis.

$$(19) H(x_n) = e^{-S(x_n)}$$

$$(20) P(x_n) = H(x_n) / \sum_{j=1}^N H(x_j)$$

For the purposes of the simulations in this article, the candidate space is always the factorial set of morpheme permutations. (Any additional candidates, even if theoretically present, would not be distinguishable by the constraints discussed here.) For example, tableau 17 shows only five selected candidates, but all $6! = 720$ permutations of the six morphemes were used as candidates in training the weights. In tableau 19 and elsewhere, all unshown candidates have effectively zero (0.00%) generated proportions.

In addition to the two most frequent outputs in 17, with RED immediately preceding or following *ka-pag*, selected other candidates shown there include (c) marginally acceptable root reduplication, (d) illicit placement of RED within *ka-pag*, and (e) illicit ordering of *ka* after *pag* rather than before it. The ten bigrams in 17 comprise the full set of bigrams encountered in the three attested outputs {*ma-RED-ka-pag-pa-ROOT*, *ma-ka-*

pag-RED-pa-ROOT, *ma-ka-pag-pa-RED-ROOT*}, following the principle that only bigrams from attested outputs need to be posited as constraints.

A tableau for the second type of variation described in §2 is given in 21. In this Tagalog example, RED can be situated in any of the three positions between the first prefix and the root, while all other morphemes are fixed in order. As before, the nine constraints represent all bigrams seen in actual outputs for this input. (A more comprehensive learning simulation combining all data in the Tagalog corpus is pursued in §5; for now, I analyze each case independently for the purposes of illustration.)

(21) Tableau illustrating variation among fixed-order affixes

			RED-pa	pa-RED	ʔi-RED	RED-ʔi	ʔi-pa	ma-ʔi	pa-√	RED-√	ma-RED
GEN'D	SCORE	CANDIDATE	13.0	12.9	11.5	11.5	10.6	8.3	7.6	6.2	5.2
a. 90.73%	46.4	ma-ʔi-RED-pa-√		*		*	*			*	*
b. 8.87%	48.8	ma-ʔi-pa-RED-√	*		*	*			*		*
c. 0.40%	51.9	ma-RED-ʔi-pa-√	*	*	*			*		*	
d. 0.00%	60.3	RED-ma-ʔi-pa-√	*	*	*	*				*	*
e. 0.00%	62.4	ma-pa-RED-ʔi-√	*		*		*	*	*	*	*

In the third type of variation in §2, two morphemes are freely ordered except in the context of some third morpheme. In Tagalog, for instance, RED can precede or follow *ka* in the prefix string *ma-ka*, but it can only precede *ka* in *ma-ka-pag*. Two tableaux are given in 22, the first for the input containing {RED, *ma*, *ka*, ROOT}, the second for {RED, *ma*, *ka*, *pag*, ROOT}. RED is generated after *ka* 2.85% of the time for the first input but never for the second.

(22) Tableaux illustrating context-sensitive variation

			pag-RED	ma-RED	ma-ka	ka-RED	RED-ka	ka-pag	pa-√	ka-√	RED-√
GEN'D	SCORE	CANDIDATE	27.8	27.5	23.2	21.1	15.6	13.1	8.5	4.9	0.0
a. 97.15%	57.4	ma-RED-ka-√			*	*		*			*
b. 2.85%	61.1	ma-ka-RED-√		*			*	*		*	
a. 61.73%	77.0	ma-RED-ka-pag-√	*		*	*				*	*
b. 38.27%	77.6	ma-ka-pag-RED-√		*		*	*		*	*	
c. 0.00%	88.9	ma-ka-RED-pag-√	*	*			*	*		*	*

The final scenario described in §2, that of free ordering among three or more morphemes, can be achieved by giving a number of bigrams the same weight. Recall Chintang, in which the prefixes *u*, *ma*, and *kha* occur freely in all permutations, while the suffixes *yokt* and *e* are fixed in order. To model these facts, the bigrams *ROOT-yokt* and *yokt-e* can be given relatively large weights (say, 20.0), while the nine remaining bigrams (three prefix-root bigrams plus six prefix-prefix bigrams) can all be assigned to the same sufficiently lower weight (say, 5.0).

5. LEARNING SIMULATIONS. In §4, I presented analyses of fragments of Tagalog verbal morphology. I now turn to a larger set of outputs annotated with corpus frequencies in order to undertake a more unified and comprehensive analysis of the prefixal morphology. As my corpus, I use the Web-based frequencies summarized in Table 2, which covers twenty-nine combinations of prefixes (as explained in §3, I ignore suffixes). For scope and realism, I also include the same twenty-nine prefix sets without RED, even though the order is always fixed in those cases. Thus, the present corpus comprises

fifty-eight prefix sets, half with RED, half without. These sets can be considered inputs. In cases of free variation, one input corresponds to multiple outputs.

Among all the outputs attested in the corpus, thirty-nine bigrams are observed; these are posited as constraints. The learner then weights the constraints so that its generated forms match its training data as closely as possible. As in §4, I use maxent learning software by Wilson and George (2008) (see Wilson 2006), but other learning algorithms for harmonic/maxent grammar could have been employed (see §6). The learned weights are given in Table 3. In training this grammar, I included all ordering permutations of each morpheme set as candidates. For example, the input set {RED, *ma*, *ka*, *pag*, *pa*, ROOT} had $6! = 720$ candidates in the training tableau, even though only three were attested. A Perl script facilitated assessing violations for each of the 4,118 candidates (all permutations of the fifty-eight input sets) against each of the thirty-nine constraints, tabulating the results in a spreadsheet formatted for the maxent software.

pa-ki	53.94	ʔi-pa	32.05	ma-pa	19.26	mag-si	3.74
ki-pag	48.59	pag-pa	31.68	si-pag	18.91	RED-pag	1.56
RED-ki	44.01	pa-ROOT	31.38	mag-ka	18.89	RED-ʔi	0.17
ma-ki	42.04	ka-RED	30.60	ma-ka	18.77	si-ROOT	0.02
ka-pag	40.90	pag-ROOT	28.80	RED-si	18.75	mag-pa	0.00
pag-RED	39.22	ki-RED	28.30	ki-ROOT	18.35	ʔi-pag	0.00
ʔi-ka	34.50	mag-RED	26.55	RED-ka	17.82	ma-pag	0.00
ʔi-RED	34.33	ka-ROOT	23.25	RED-pa	13.55	RED-ma	0.00
pa-RED	34.29	ma-RED	23.08	ma-ʔi	13.15	ma-ROOT	0.00
ʔi-ROOT	33.59	si-RED	22.68	RED-ROOT	11.23		

TABLE 3. Constraint weights after learning.

Various criteria can be used to assess how well the learned grammar matches its training data. The mean percentage error for the ninety-seven RED positions given in Table 2 is 1.16%. The weighted mean error, in which observed/generated differences are given weight in proportion to the frequency of the input, is 0.13%. Pearson's correlation coefficient is 0.9954 for all forms observed and/or generated with greater than 0.0% incidence. Finally, the three maximum errors are 29.4%, 17.9%, and 11.5%, respectively, all for the relatively poorly attested input {RED, *ʔi*, *pa*, *ki*, ROOT}. While it is difficult to say what constitutes a 'good' or 'bad' fit in absolute terms, these values become more meaningful when theories are compared. To this end, I also trained a grammar of precedence constraints (X ... Y: 'assess a violation iff X does not precede Y'; see §7) on the same data. Forty-three constraints were posited, one for each precedence pair observed in the training data. Despite its greater number of free parameters (i.e. constraints), the fit of the optimal precedence grammar is substantially worse on all criteria. The unweighted and weighted mean percentage errors for the ninety-seven RED positions are 13.55% and 7.09%, respectively, over ten times as great as for the bigram grammar. The correlation is 0.8195. The three maximum errors are 83.8%, 77.1%, and 65.8%.

In this section, I trained the grammar on all available corpus data. It is also possible to hold back data from the learner and then test the resulting grammar on unseen data to see how well it differentiates actually grammatical suggestions from ungrammatical ones, as a form of cross-validation. I pursue one such simulation in §6.

6. MORPHOTACTIC EXTENSION. What are possible and impossible patterns of free variation? Why does variation exist at all? In this section, I intend to begin to answer these questions (see also §10) by describing an emergent property of morphotactic constraint systems I call morphotactic extension, which can be defined as the analogical extension of an ordering relationship among morphemes to a context in which it was not previ-

ously encountered. My claim, in other words, is that affix order is subject to analogy, and that this analogy can help explain the diachronic origin and synchronic stability of morphotactic variation.

To illustrate this principle with a simple Tagalog example, consider the prefix strings *ma-ka*, *pag-pa*, and *ma-ka-pag-pa*. In Modern Tagalog, RED is relatively stable in the first two (97.2% *ma-RED-ka* and 100.0% *pag-RED-pa*) but varies between two positions in the third (74.6% *ma-RED-ka-pag-pa* ~ 25.4% *ma-ka-pag-RED-pa*). But imagine, for the sake of illustration, that we are dealing with a hypothetical stage of pre-Tagalog in which RED's placement is always categorical, for example, a second-syllable reduplicant (cf. Inkelas & Zoll 2005). Thus, our hypothetical pre-Tagalog learner encounters only *ma-RED-ka*, *pag-RED-pa*, and *ma-RED-ka-pag-pa*. Furthermore, as in Modern Tagalog, we can assume that *ma-ka* (+RED) and *pag-pa* (+RED) are vastly more frequent than *ma-ka-pag-pa* (+RED) (in Table 2, they are 139 and twenty times as frequent, respectively).

Bigram morphotactics predicts that until a late stage of learning, the learner will substantially overgenerate RED between *pag* and *pa* in *ma-ka-pag-pa*, creating variation where there was none in its training data (in this hypothetical historical scenario). Figure 3 schematizes the extension of RED from between *pag* and *pa* in *pag-RED-pa* to that same position in *ma-ka-pag-RED-pa*. In this case, the learner posits and gives some weight to the constraints *pag-RED* and *RED-pa* when it sees RED in those contexts in the prefix string *pag-RED-pa*. In doing so, however, it initially also (over)generates RED between *pag* and *pa* in the longer prefix string *ma-ka-pag-pa*, even though it has not seen RED in that position in that string. The key to morphotactic extension is the locality of the constraints: the learner initially anticipates the position between *pag* and *pa* to be felicitous based on its knowledge, encoded in bigram constraints and their weights, that RED occurs after *pag* and before *pa* in other, more common prefix strings.

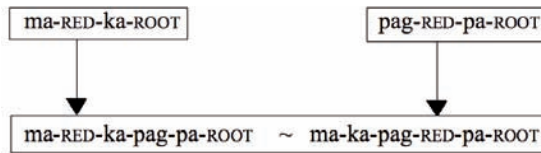


FIGURE 3. Simple illustration of morphotactic extension.

Of course, the learner would also encounter many other prefix strings containing these affixes, so in order to show more rigorously how extension exerts its effects during learning, a more comprehensive simulation is in order. As my training corpus, I use the Web-based corpus (Table 2), only now with all variants suppressed. Only the 'best' (most frequent) output is represented for each input. In order to model the timecourse of learning, I employ a gradual learning algorithm for harmonic grammar described by Boersma and Pater (2008) and Pater (2009:19), which is based on Rosenblatt's (1958) Perceptron and only slightly modified from the gradual learning algorithm for stochastic optimality theory (Boersma 1997, Boersma & Hayes 2001). See also Jäger 2007 on an online maxent learner.

A babblar, written in Perl, feeds Tagalog verb tokens to the learner one at a time, randomly selected from the training corpus in proportion to their frequencies. The learner, again implemented in Perl by the author, takes two steps for each token. First, if it sees any bigrams that are not yet part of its constraint set, it posits them as constraints, initially weighted zero. (I assume that the learner starts out already mature enough to parse affixes; on parsing, see Goldsmith 2001, Baroni 2003, Poon et al. 2009, and many oth-

ers. Here I treat only the evolution of the productive grammar of ordering.)⁴ Second, the learner undergoes error-driven reweighting. Learning is error-driven in the sense that the learner takes the set of morphemes in the observed token and compares its generated output for that set of morphemes. If they are the same, nothing further happens. If there is a discrepancy, however, an error has occurred, to which the learner responds by adjusting its constraint weights, specifically, incrementing any constraints preferring the observed form and decrementing any preferring the error. Finally, variation between outputs is achieved by adding random (Gaussian) noise to each weight at each evaluation, causing closely weighted constraints to sometimes trade off dominance. Additional details about these processes are not important here and can be found in the cited references.

Crucial for the present purposes is that if we stop this learner at a sufficiently early (but not, as is explained below, overly early) stage of learning and test it on all inputs, it generates significant variation for some inputs, despite having observed none. Moreover, this variation resembles the variation found in the actual Tagalog corpus, to which the learner did not have access.⁵ For example, after training on 30,000 tokens from the best-only corpus, the learner generates RED as in Table 4, which is laid out like Table 2. Percentages from the actual corpus (Table 2) are repeated here to the left of each slash; percentages generated by the learner are to the right. For each RED position seen by the learner, the number of tokens seen is given in parentheses (these do not add up to 30,000 because Table 4 shows only inputs containing RED and at least two other prefixes). Since the training corpus contains no variation, the learner never sees more than one RED position used for any particular input. For a few prefix strings (e.g. *?i-pa-ki*), the learner did not see any tokens with RED.

Despite its categorical training data, the learner generates variation for most inputs in Table 4. For example, for the prefix string *ma-ka-pag-pa*, the learner only saw RED between the first two prefixes (twenty-seven tokens). But it generates RED in two positions (72.1% *ma-RED-ka-pag-pa* ~ 26.5% *ma-ka-pag-RED-pa*), poorly matching its training data, but uncannily approximating the real Tagalog corpus (i.e. 74.6% *ma-RED-ka-pag-pa* ~ 25.4% *ma-ka-pag-RED-pa*). As outlined above, this prediction is driven by extension from other prefix strings (e.g. 728 tokens of *pag-RED-pa*). The learner has not yet seen enough tokens of *ma-ka-pag-pa* (with RED always in the second position) to ‘unlearn’ this possibility, as it eventually would if learning were allowed to proceed. Because short prefix strings vastly outnumber longer ones, the learner’s early treatment of longer strings is based more on extrapolation from short strings than on actual experience with longer ones (cf. Elman 1993 on how ‘starting small’ can facilitate learning under certain conditions).

The learner also correctly anticipates, for the most part, where extension is blocked. Compare, for instance, *ma-ka-pag* and *ma-ki-pag* (Table 5). For both inputs, the learner

⁴ A referee asks whether starting out with SCOPE hardwired and positing morphotactics during learning predicts that scopal ordering would be default in early morphological acquisition. This proposal does not necessarily make any predictions one way or the other on this issue. First, it is possible that scope starts out with a weight of zero, in which case it would be inert initially. More importantly, in actual acquisition, morphologically complex forms are learned and used before being fully parsed (Courtney & Saville-Troike 2002); indeed, this is a prerequisite to learning morphology. Thus, the learner might use both scopal and templatic orderings simply because he or she has memorized those tokens.

⁵ I wish to thank Joe Pater for drawing my attention to the possibility of investigating gradual learning effects in the context of my discussion of morphotactic extension. I had previously demonstrated the predictability of marked variants using a smoothing factor (Gaussian prior) in maxent learning.

BASE PREFIX STRING	POS. 0	POS. 1	POS. 2	POS. 3	POS. 4
?i-ka-	0/0	28.5/32.1	71.5/68.0 (186)		
?i-pa-	0/0	83.9/97.9 (685)	16.1/2.1		
?i-pag-	0/0	1.5/0.1	98.5/99.9 (241)		
ka-pag-	37.4/0	0/4.7	62.6/95.3 (13)		
mag-ka-	0/0	99.9/99.9 (474)	0.1/0.1		
mag-pa-	0/0	100/100 (842)	0/0		
mag-si-	0/0	99.9/97.9 (99)	0.1/2.1		
ma-?i-	0/0	13.5/13.3	86.5/86.8 (482)		
ma-ka-	0/0	97.2/98.1 (5,112)	2.8/1.9		
ma-ki-	0/0	95.5/100 (234)	4.5/0		
ma-pa-	0/0	96.0/100 (27)	4.0/0		
ma-pag-	0/0	5.7/29.1	94.3/70.8 (2)		
pag-pa-	0/0	100/95.3 (728)	0/4.7		
pa-ki-	0.6/0	92.2/79.7 (96)	7.2/20.3		
?i-ka-pag-	0/0	2.8/4.3	0.3/7.7	96.9/88.1	
?i-pag-pa-	0/0	6.8/0.1	93.2/99.9 (39)	0/0	
?i-pa-ki-	0/0	17.9/14.0	62.6/69.9	19.5/16.2	
mag-si-pag-	0/0	100/38.7 (4)	0/0.8	0/60.6	
ma-?i-pa-	0/0	0.3/15.3	90.9/81.6 (60)	8.8/3.1	
ma-?i-pag-	0/0	0.1/0	1.5/0.7	98.4/99.3 (14)	
ma-ka-pag-	0/0	61.8/69.2 (103)	0/1.6	38.2/29.3	
ma-ki-pag-	0/0	99.9/97.1 (27)	0/0	0.1/2.9	
?i-ka-pag-pa-	0/0	1.5/4.0	0.8/8.5	97.7/87.5	0/0
ma-?i-pag-pa-	0/0	0/0.1	4.0/0.6	96.0/99.4 (6)	0/0
ma-ka-pag-pa-	0/0	74.6/72.1 (27)	0/1.4	25.4/26.5	0/0

TABLE 4. Learner's predictions after training on filtered corpus.

only saw RED immediately following *ma*. For *ma-ka-pag*, it additionally extends to the position between *pag* and the root. But it does not (more than slightly) extend to the same position in *ma-ki-pag*. This is because, based on its training data, it has given RED-*ki* more weight than RED-*ka* (28.3 and 21.8, respectively). (The learner erroneously extends RED to the preroot position in *mag-si-pag* because it has not yet seen enough tokens with *si* to know that *si* is more like *ki* than *ka* when it comes to blocking extension.)

OUTPUT	ACTUAL CORPUS	SEEN BY LEARNER	GENERATED BY LEARNER
ma-RED-ka-pag-ROOT	61.8%	100%	69.2%
ma-ka-pag-RED-ROOT	38.2%	0%	29.3%
ma-RED-ki-pag-ROOT	99.9%	100%	97.1%
ma-ki-pag-RED-ROOT	0.1%	0%	2.9%

TABLE 5. A closer look at some of the learner's predictions.

Thus, on average, the learner's (over)generation of variation is localized to positions where variation is exhibited in the actual Tagalog corpus. Comparing actual to generated percentages in Table 4, the overall weighted mean error (as in §5) is 2.9%. If the learner had matched its training data perfectly, this error would have been 4.6%, almost twice as great. Indeed, if trained for long enough on the best-only corpus, the learner eventually (after around a million tokens) converges on a virtually categorical grammar. Not surprisingly, the observed-to-generated fit increases monotonically as a function of training. More interesting, however, is that the best-only-trained learner's fit to the unobserved actual corpus peaks at ~30,000 tokens and declines from then on. The gener-

ated-to-actual fit (by ‘fit’ I mean 100% minus the weighted mean error) is plotted against time (log tokens babbled) in Figure 4. The dotted line at 95.4% indicates the fit between the training and actual corpora. The generated-to-actual fit converges to this line at asymptote as the learner perfects its match to the training corpus.

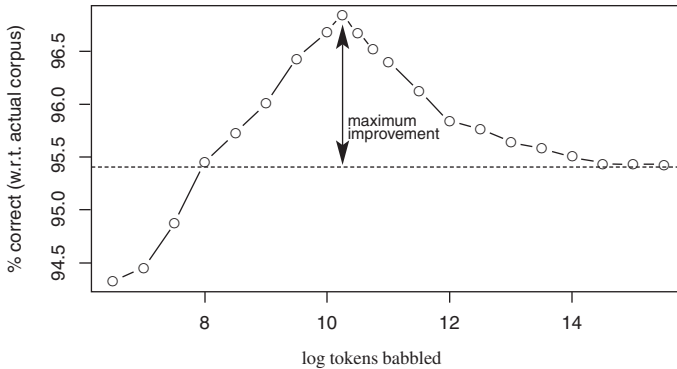


FIGURE 4. Learner’s improvement over its impoverished training corpus over time.

To conclude, in the previous section I showed that bigram morphotactics enables the learner to closely match its training data when trained on the whole corpus. In this section I showed that even when it is trained on a filtered corpus of invariable (most frequent output only) Tagalog data, the semi-mature learner still correctly predicts to a large extent which inputs should exhibit variation and how that variation should manifest. It follows that under bigram morphotactics, the marked variants are not arbitrary; rather, they emerge naturally from extension from unmarked outputs. In Tagalog, the invariable corpus is actually harder for the learner to master than the variable corpus, as evidenced by the learner’s initial skewing toward the latter. Diachronically, morphotactic extension provides an analogical explanation for the innovation of variation in categorical systems with certain characteristics. Synchronically, it motivates the stability of marked variants in the productive morphology. Consider a rare but grammatical position for RED such as *ma-RED-?i-pa-ROOT* (0.3% incidence, but explicitly identified as grammatical in S&O:362). This form is rare enough that a young learner might never hear it. Under bigram morphotactics, however, this position, unlike certain other unseen positions, receives support from extension.

Finally, morphotactic extension constrains the typology by rendering certain types of systems unlearnable. Imagine, for instance, a hypothetical version of Tagalog in which RED is placed between *pag* and the root unless the suffix *an* is present, in which case RED must precede *pag*. The bigram morphotactic learner can never converge on this system; it can only learn variation for both inputs, unavoidably extending to both. Extension can be blocked only when the blocking morpheme is adjacent to the morpheme whose extension is being blocked (as with *ki* blocking the extension of RED in Table 5, in which the bigram RED-*ki* is able to ‘attract’ RED away from what would otherwise be an eligible position; there is no constraint RED-*an* or *an*-RED to play the role of attractor in the present case). I return to restrictiveness in §10.

7. AGAINST OTHER MORPHOTACTIC THEORIES. I showed in §4 that bigram constraints can model the types of variation described in §2. I now argue that other possible morphotactic formalisms, namely, morpheme alignment, precedence constraints, and a sin-

gle big template, undergenerate these scenarios. I also make some preliminary remarks against the morphosyntactic accounts of RED variation proposed by Rackowski (1999), Mercado (2007), and Skinner (2008) before going on to make more general arguments against serial rule-based approaches to affix order in §8.

First, an approach to morphotactics sometimes used in OT is that of morpheme alignment (e.g. Hargus & Tuttle 1997, Trommer 2003, Jaker 2006). An alignment constraint conforming to the schema proposed by McCarthy and Prince (1994) can be used to enforce the proximity of a morpheme to the edge of a phonological constituent such as the prosodic word (PWD). For instance, ALIGN(*its*, left, PWD, left), which I abbreviate ‘its-L’ in 23, is violated by the number of morphemes (or some other unit) that intervene between the left edges of the PWD and *its*. This constraint therefore pressures *its* to be word-initial. Constraints specifying other morphemes can be formulated on this pattern and ranked against each other, as in 23. This ranking guarantees that *its* always precedes *il* in the output, regardless of scope.

(23) Tableau illustrating morpheme-alignment constraints

/[[[ROOT]il]its]/	ROOT-L	its-L	il-L
a. \mathcal{C} ROOT-its-il		*	**
b. ROOT-il-its		**!	*
c. its-ROOT-il	*!		**

Alignment cannot generate the first three scenarios in §2. It cannot model an affix varying freely between nonadjacent positions, since that would require the alignment constraint of the variable affix to float freely between nonadjacent positions in the hierarchy over a set of fixed constraints: {RED-L >> *ka*-L >> *pag*-L; *ka*-L >> *pag*-L >> RED-L} (for RED-*ka*-*pag* ~ *ka*-*pag*-RED). A partially ordered grammar (as in Anttila 1997a,b) could specify *ka*-L >> *pag*-L, leaving RED-L unranked with respect to *ka*-L and *pag*-L, but that grammar would also generate **ka*-RED-*pag*. A similar problem obtains for statistical models such as stochastic OT (Boersma 1997, Boersma & Hayes 2001), maxent grammar (see §4), and noisy harmonic grammar (see §4). If only morpheme alignment to the word edge is employed, the pre-*ka* and post-*pag* positions cannot be generated to the exclusion of the position between *ka* and *pag*.

Alignment also cannot model the scenario in which an affix is freely situated among a fixed sequence of morphemes, as in 10. Under partial ordering, RED-L must occupy the same stratum as both *ʔi*-L and *pa*-L, since RED can precede or follow each of these prefixes. This entails that *ʔi*, *pa*, and RED all be freely ordered. A stochastic implementation would require the probability distribution of RED-L to have a greater standard deviation than that of *ʔi*-L and *pa*-L, so that RED-L might fall anywhere around these two constraints, which remain narrowly pinned down. But deviation is not a free parameter in the cited models (cf. Reynolds 1994, Nagy & Reynolds 1997). Finally, alignment cannot implement context-sensitive reorderability, as in 11. Because *ka* and RED are swappable in *ma*-*ka*-ROOT, the variable ranking *ka*-L ~ RED-L is necessary. This variable ranking incorrectly requires that *ka* and RED be equally swappable in all contexts. For additional arguments against ordering affixes by alignment, see Rackowski 1999 and §9 below.

Another morphotactic formalism is the precedence constraint, as found in Muysken 1981b:266ff., Paster 2006b:184, and Caballero 2011, for example, ‘C > A: Causative precedes Applicative’ (Caballero). A precedence constraint can be considered the same as a bigram except that a bigram evaluates only adjacent pairs of morphemes, while a prece-

dence constraint evaluates all pairs, regardless of adjacency. Precedence constraints cannot implement variation between nonadjacent positions, for example, Tagalog RED-*ka-pag* ~ *ka-pag*-RED. To select the RED-*ka-pag* variant, the following ranking is needed: {RED > *ka*, RED > *pag*} >> {*ka* > RED, *pag* > RED}. For the *ka-pag*-RED variant, the two strata must be swapped wholesale: {*ka* > RED, *pag* > RED} >> {RED > *ka*, RED > *pag*}. Such contingency in ranking variation is not possible in the cited frameworks. Moreover, precedence cannot model context-sensitivity in affix ordering, for example, Chichewa *an* 'reciprocal' and *its* 'causative' being swappable except following *il* 'applicative'. *an* > *its* and *its* > *an* must be freely ranked to motivate the variation, but then there is no way to make the variation contingent on the presence of *il* earlier in the word.

Other theorists employ a monolithic template (see Stump 2006 and references therein) to enforce arbitrary ordering restrictions, for example, Hyman's (2003) violable 'CARP' constraint: 'causative > applicative > reciprocal > passive'. As Paster (2006b:184) points out, such templates cannot account for cases such as 8, in which two affixes are freely ordered for both scopes. Scenarios with multiple free variants (e.g. the third and fourth types of variation described in §2) make the need for multiple interacting morphotactics even more apparent. See §9 for additional arguments against templatic theories.

Finally, I briefly address two morphosyntactic accounts of affix order variation in Tagalog. Rackowski (1999) proposes a DISTRIBUTED MORPHOLOGY (Halle 1990, Halle & Marantz 1993) account of the variable placement of RED involving optional morphological movement, which she likens to scrambling (see also Vaux 2002). RED is generated next to the root and optionally raises in the morphology to adjoin to any vP head within its phase (see Chomsky 2001 on phases). The phase is taken to be closed by the topic marker, for example, *ʔi* 'object topic' or *m* 'actor topic'. This phase-bounded movement ostensibly explains why RED is usually ungrammatical preceding a topic marker, but grammatical in various positions after it. Mercado (2007) endorses the phase-boundedness of Rackowski's analysis. Nevertheless, observing that Rackowski's low generation of RED under morphemes such as the causative is incompatible with the fact that RED invariably scopes over such morphemes (see also Travis 2007), Mercado eliminates affix movement, instead envisioning RED as a multipositional head. As such, RED is base-generated simultaneously in all intermorphemic (pre-vP) positions in its phase, including next to the root. In the morphology, a constraint of unique instantiation (Noyer 1993) ensures that only one position of this head is spelled out. Finally, Skinner (2008) argues that RED is generated high in the structure, just inside the topic marker, and optionally lowers to any preroot position.

These phase-bounded analyses greatly overgenerate positions for RED. As described in §3, in a verb such as *ma-ka-pag-pa*-ROOT, RED is virtually always (over 99.9% of the time) placed before *ka* or *pa*. Yet the phase-based analyses fail to differentiate among any of the four positions after *ma*, merely licensing them all as grammatical, even the positions (e.g. *pre-pag* here) that the grammarians (e.g. S&O) identify as ungrammatical. Moreover, these analyses fail to capture the crucial role context plays in determining whether a particular position is available. For instance, RED is sometimes acceptable between *pag* and the root and sometimes not, depending on which other morphemes are present. In *mag-si-pag*-ROOT, RED is never found between *pag* and ROOT, while in *ma-ka-pag*-ROOT, RED is common in that position (38.2% incidence). Finally, these analyses undergenerate outside the phase. For example, in *ma-ʔi*-ROOT, RED precedes *ʔi* 12.8% of the time in the corpus, which is also regarded as a grammatical position in S&O (362).

But *?i* is a topic marker, so the phase-bounded analyses predict this position to be completely ungrammatical.

8. PARALLEL VS. SERIAL THEORIES OF AFFIX ORDERING. Language-specific rules of post-syntactic movement or dislocation are sometimes invoked to explain arbitrary ordering restrictions (see e.g. Embick & Noyer 2001, Embick 2007 on these operations in distributed morphology). Movement and dislocation are operations in the sense that they are structure-altering rules applying in a serial derivation, each rule being blind to any subsequent rule applications (Chomsky 1995, 2001, Embick 2007:329, 2008). A constraint-based theory of morphotactics, by contrast, selects the optimal output(s) from a set of candidates evaluated in parallel by a system of ranked or weighted constraints (on the general theory of constraint interaction, see Prince & Smolensky 2004 [1993] and Smolensky & Legendre 2006).

In this section I briefly address some general differences between serial-operational and parallel theories of affix ordering. Consider the fixed ordering of the causative and applicative suffixes in Chichewa. When these are the only two derivational morphemes in the verb, the causative *its* must always precede the applicative *il*, regardless of which scopes over the other (Hyman 2003). For an applicativized causative, this order is scopal; but for a causativized applicative, it is counterscopal. In a parallel morphotactic theory, a constraint such as CAUS-APPL could be active, dominating any constraints preferring the other order (e.g. other morphotactics and/or, as Hyman proposes, a MIRROR constraint).

In an operational framework such as Embick and Noyer's (2001), the counterscopal ordering could be achieved by movement. Rackowski (1999) suggests that this case involves lowering, perhaps as sketched in Figure 5a, in which the causative head lowers to adjoin to the applicative head. Local dislocation could not be invoked here because the causative skips over intervening morphemes in more complex verbs such as the causativized reciprocalized applicative in Figure 5b, though it can optionally remain in its base position in such verbs (data from Hyman 2003:273). (It is possible that a more powerful version of lowering than what Embick and Noyer (2001) assume is required for this case, such as the version advocated by Skinner 2008.) These rules might be formalized somewhat differently by different authors, but here I intend only to sketch some general differences in the perspectives and predictions of constraint- vs. rule-based frameworks, since there undoubtedly are such differences that transcend certain details of implementation (see also Embick 2008:§1).

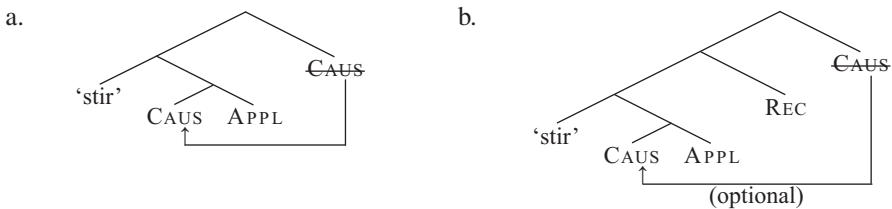


FIGURE 5. Sketches of causative lowering in Chichewa.

In the operational framework, it would be possible to add a second rule to Chichewa moving APPL to CAUS in applicativized causative verbs (not shown). Both of these rules together in the language would make for a hypothetical version of Chichewa (call it Chichewa') in which both causativized applicatives and applicativized causatives sur-

faced in counterscopal order.⁶ I call this hypothetical phenomenon SCOPAL METATHESIS, in the sense that the two suffixes always surface in counterscopal order. No such language is attested.

Constraint-based morphotactics properly rules out scopal metathesis. Affixes can be realized in counterscopal order only when that order is morphotactically motivated, that is, when the reordering (*vis-à-vis* scope) results in more regularity in order across all outputs. In a system of morphotactic and SCOPE constraints, the complete typology of possible interactions between two affixes was given in 10. This typology includes fixed ordering (morphotactics dominates), scopal ordering (SCOPE dominates), asymmetric compositionality (morphotactics and SCOPE are variably ranked), and free ordering (morphotactic constraints are variably ranked). No ranking can generate scopal metathesis. (Free ordering is not scopal metathesis, since there is no requirement that the affixes surface in counterscopal as opposed to scopal order.)

This fact about the typology extends to all constraint-based morphotactic theories, even if multiple morphotactics interact, as I maintain. It is guaranteed by Moreton's CHARACTERIZATION THEOREM of violable constraint evaluation (1999), according to which deviations from faithfulness (analogous to scope) are possible only if they reduce markedness (analogous to morphotactics). This theorem does not hold of operational approaches; for example, one could posit a rule of symmetrical metathesis, $V_1V_2 \rightarrow V_2V_1$, that converts /oa/ into [ao] and /ao/ into [oa] (Moreton 1999:17). Symmetrical metathesis is analogous to scopal metathesis, in which the surface order of two morphemes is always the reverse of their base order. Just as constraint-based phonology with markedness and faithfulness constraints cannot generate symmetrical metathesis, constraint-based morphology with morphotactic and scope constraints cannot generate scopal metathesis.

A second unattested scenario that derivations can generate but morphotactic constraints cannot is the case of asymmetric compositionality (see 7 and 10) in which the fixed order is counterscopal rather than scopal. This could be achieved by having optional movement or dislocation for one scope (as required for the attested pattern) as well as obligatory movement or dislocation for the other scope.

In sum, constraint-based morphotactics predicts that ordering will be either scopally licensed (iff a SCOPE constraint is active) or else insensitive to syntax/semantics, being driven entirely by the active morphotactics. An operational framework, by contrast, predicts (without limiting stipulations) that orderings that do not directly reflect scope could nevertheless be determined by scopal information, as in the case of scopal metathesis, in which the surface order is never faithful to scope but always determined by it. This extra power afforded by derivations is apparently unnecessary.

⁶ The argument here does not depend on the rules being lowering as opposed to raising, dislocation, phrasal movement, or any other operation having the effect of rendering the scope [[[root]CAUS]APPL] as [...ROOT-APPL-CAUS...]. Furthermore, the additional rule cannot be ruled out by universal stipulation. As we saw in Pulaar (8), the applicative optionally precedes the causative in applicativized causatives, so this rule would be needed in that case. A referee points out that my argument would not apply if the rule-based formalism were such that the two rules could only be formulated such that one would obligatorily feed the other. This is not a property of Embick and Noyer's (2001) framework, in which complex heads, once created by adjunction, can be moved as units (e.g. 2001:572). In the case of Chichewa', if we take CAUS lowering as the first rule, CAUS would first lower to APPL to form a complex APPL head. Then, for feeding to be obligatory, the second rule would have to move APPL within its complex head, taking the APPL morpheme and lowering it to CAUS, which had previously adjoined it.

9. NONTRANSITIVITY. By NONTRANSITIVITY in morpheme order, I refer to the situation in which the ordering restrictions cannot be expressed collectively as a single total or partial order over morphemes. Bigram morphotactics, unlike certain other theories, predicts nontransitivity (see also Anderson 1986). Consider such a case in the abstract: (a) morpheme X must precede Y, (b) Y must precede Z, but (c) X must follow (or optionally follows) Z. This scenario cannot be implemented in a framework in which affix ordering is required to be transitive, for example, alignment (see §7). Under alignment, (a) requires X-L [= *Align*(X,left,PWd,left)] >> Y-L and (b) requires Y-L >> Z-L. Because of ranking transitivity, (a) and (b) entail X-L >> Y-L >> Z-L, which contradicts (c), according to which X always or sometimes follows Z. Nontransitivity is also problematic for theories in which affix order falls out from the association of each affix with a coherent level, stratum, or position class (e.g. Kiparsky 1982, 2000, Mohanan 1986, Inkelas 1993).

In doubly derived Chumbivilcas Quechua verbs, for instance, *ri* ‘inchoative’ can only precede *sch*i ‘assistive’ and *sch*i can only precede *na* ‘reciprocal’ (Muysken 1988:263). If transitivity held, *ri* would have to precede *na*. But, in fact, *na-ri* is the only acceptable order, violating transitivity. As a second example, this time involving variation, *paya* ‘frequentative’ must precede *ru* ‘exhortative’ and *ru* must precede *sch*i. But *paya* and *sch*i can occur in either order, again violating transitivity (this is true regardless of whether the variation of *paya* and *sch*i is free or correlated with scope). Bigram analyses of these two cases are given in 24 and 25, respectively.

(24) Example of fixed nontransitivity in Chumbivilcas Quechua

	ri-schi	schi-na	na-ri
a. ☞ -ri-schi-		*	
b. -schi-ri-	*!	*	
a. ☞ -schi-na-			*
b. -na-schi-		*!	*
a. ☞ -na-ri-	*		
b. -ri-na-	*		*!

(25) Example of variable nontransitivity in Chumbivilcas Quechua

	paya-ru	ru-schi	paya-schi	schi-paya
a. ☞ -paya-ru-		*	*	
b. -ru-paya-	*!	*	*	
a. ☞ -ru-schi-				*
b. -schi-ru-		*!		*
a. ☞ -paya-schi-	*			*(!)
b. ☞ -schi-paya-	*		*(!)	

These are by no means isolated cases. Muysken (1988), for one, systematically investigated all 462 (22² – 22) pairwise-distinct binary combinations of twenty-two derivational suffixes in Chumbivilcas Quechua (Cuzco region, Peru), documenting each as ‘+’ (acceptable order), ‘–’ (never acceptable), or ‘?’ (indeterminate or inconsistent data). Putting aside ‘?’ cases, there are forty-four distinct cases of nontransitivity evident in this chart. Five of these are summarized in Table 6, with ‘>’ indicating a fixed ordering and ‘~’ variation. Van de Kerke (1996:43) prepared a similar chart for sixteen suffixes in Tarata Quechua, a South Bolivian variety. I found fourteen nontransitive sets

in van de Kerke's data, five being shown in Table 6.⁷ Buell and colleagues (2008:5) discuss a comparable case in Wolof in which *loo* 'causative' > *al* 'benefactive', *al* > *e* 'instrumental', but *e* > *loo*. Grimes (1983) describes another case in Tucano, and Spencer (1991:210ff.) discusses one in Navajo (see also Ackema & Neeleman 2005:306). Bobaljik (2000) treats similar cases in adverb ordering.

CHUMBIVILCAS	(a) <i>chi</i> > <i>sch</i> i	(b) <i>sch</i> i > <i>na</i>	(c) <i>chi</i> ~ <i>na</i>
	(a) <i>naya</i> > <i>puna</i>	(b) <i>puna</i> > <i>ri</i>	(c) <i>naya</i> ~ <i>ri</i>
	(a) <i>paku</i> > <i>yu</i>	(b) <i>yu</i> > <i>chi</i>	(c) <i>paku</i> ~ <i>chi</i>
	(a) <i>puna</i> > <i>ru</i>	(b) <i>ru</i> > <i>sch</i> i	(c) <i>sch</i> i > <i>puna</i>
	(a) <i>yu</i> > <i>sch</i> i	(b) <i>sch</i> i > <i>puna</i>	(c) <i>puna</i> > <i>yu</i>
TARATA	(a) <i>k</i> ipa > <i>naya</i>	(b) <i>naya</i> > <i>ra</i>	(c) <i>k</i> ipa ~ <i>ra</i>
	(a) <i>naya</i> > <i>ra</i>	(b) <i>ra</i> > <i>chi</i>	(c) <i>naya</i> ~ <i>chi</i>
	(a) <i>paya</i> > <i>raya</i>	(b) <i>raya</i> > <i>ri</i>	(c) <i>paya</i> ~ <i>ri</i>
	(a) <i>naya</i> > <i>ra</i>	(b) <i>ra</i> > <i>mu</i>	(c) <i>naya</i> ~ <i>mu</i>
	(a) <i>naya</i> > <i>ra</i>	(b) <i>ra</i> > <i>pu</i>	(c) <i>naya</i> ~ <i>pu</i>

TABLE 6. Ten examples of nontransitivity from Chumbivilcas and Tarata Quechua.

What happens when all three suffixes in one of these pairwise nontransitive sets cooccur? Consider the Chumbivilcas nontransitive set: *la* 'just' > *n* '3', *n* > *kuna* 'PL', *kuna* > *la*. Muysken reports that for a triply affixed noun containing *la*, *n*, and *kuna*, free variation obtains between *-la-n-kuna* and *-n-kuna-la* (1981b:295; §2 above). These four tableaux (three pairs plus one triplet) are given in 26.

(26) Nontransitivity in pairs and triplets of suffixes in Chumbivilcas Quechua

	n-kuna	la-n	kuna-la
a. ☞ -la-n	*		
b. -n-la	*	*!	
a. ☞ -n-kuna			*
b. -kuna-n	*!		*
a. ☞ -kuna-la		*	
b. -la-kuna		*	*!
a. ☞ -la-n-kuna			*(!)
b. ☞ -n-kuna-la		*(!)	
c. -n-la-kuna	*!	*	*
d. -la-kuna-n	*!	*	*
e. -kuna-n-la	*!	*	*
f. -kuna-la-n	*!		

A different, context-sensitive type of nontransitivity is exhibited when the ordering of a pair of affixes depends on whether or not some third affix is present. In Huave, for instance, reflexive *ay* can usually precede only first-person *Vs* (27a) (Stairs & Hollenbach 1981, Embick & Noyer 2001:576). But in 27b, in the context of plural *on*, *ay* must follow *Vs*, contradicting the *ay-Vs* ordering implied by 27a.

⁷ Morpheme glosses are as follows. Chumbivilcas: *chi* 'causative', *na* 'reciprocal for transitives', *naya* 'desiderative', *paku* 'for someone else's benefit', *puna* 'reciprocal for intransitives', *ri* 'inchoative', *ru* 'exhortative', *sch*i 'assistive', *yu* 'augmentative'. Tarata: *chi* 'causative', *k*ipa 'repetitive', *naya* 'desiderative', *pu* 'benefactive', *ra* 'distributive', *ri* 'inceptive', *raya* 'stative'.

- (27) a. t-e-kohč-ay-os, *t-e-kohč-as-ay
 PAST-THEME-cut-REFL-1 PAST-THEME-cut-1-REFL
 ‘I cut (PAST) myself’
 b. t-e-kohč-as-ay-on, *t-e-kohč-ay-as-on
 PAST-THEME-cut-1-REFL-PL PAST-THEME-cut-REFL-1-PL
 ‘we cut (PAST) ourselves’

A similar example can be found in Lithuanian (Senn 1966, Nevis & Joseph 1993, Embick & Noyer 2001:578), in which *si* ‘reflexive’ is a suffix, unless one or more of a certain class of prefixes is present, in which case *si* occurs as the second prefix, as in 28.

- (28) a. laikaũ-si
 maintain-REFL
 ‘I get along’
 b. iš-si-laikaũ
 PREVERB-REFL-maintain
 ‘I hold my stand’

Bigram analyses of 27 and 28 are sketched in 29 and 30, respectively.

(29) Nontransitivity in Huave

	ay-on	Vs-on	ay-Vs	Vs-ay
a. ☞ -ay-os	*	*		*
b. -as-ay	*	*	*!	
a. ☞ -as-ay-on		*	*	
b. -ay-as-on	*!			*

(30) Nontransitivity in Lithuanian

	iš-si	iš-ROOT	ROOT-si	si-ROOT
a. ☞ laikaũ-si				*
b. si-laikaũ			*!	
a. ☞ iš-si-laikaũ		*	*	
b. iš-laikaũ-si	*!			*

In short, bigram morphotactics can implement the sorts of nontransitive ordering restrictions instantiated in natural languages. Still, one might justifiably wonder why nontransitivity is as uncommon as it is, given how many rankings/weightings generate it on this proposal. I return to this question at the end of the next section, which addresses the restrictiveness of the theory.

10. ON THE RESTRICTIVENESS OF BIGRAM CONSTRAINTS. Bigram morphotactics is a restrictive theory of affix order variation. Unlike a brute-force approach such as representing possible orderings as a finite-state automaton (e.g. the directed graphs used for illustrative purposes in Jurafsky & Martin 2000:§3, Plag & Baayen 2009),⁸ there are many unattested but logically possible types of affix order variation that constraint-based morphotactics cannot generate. Recall, for instance, the attested pattern of context-sensitive variation described in 11, which might be schematized /A,B,C/ → {A-B-C, B-A-C}; /A,B,D/ → {A-B-D} (*B-A-D). In this case, A and B are freely ordered, except when D follows them. This pattern can be modeled by a finite-state au-

⁸ Finite-state machines are not ‘brute force’ in general, but if we assume that a language has finitely many affixes, then there is no set of permutations that a finite-state machine cannot generate.

tomaton (Figure 6), which accepts only the strings A-B-C, B-A-C, and A-B-D. State 0 is the start state and 5 is the final state (graphs from Graphviz; Gansner & North 1999).

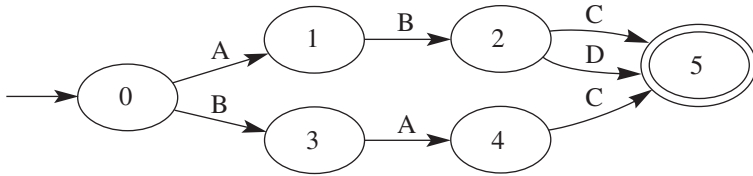


FIGURE 6. Finite-state automaton for morpheme order.

To model gradient data, a probability could be assigned to each edge. To restrict the machine to be sensitive to the input, a transducer could be employed, adding an input morpheme to each edge in Fig. 6, as in Figure 7.

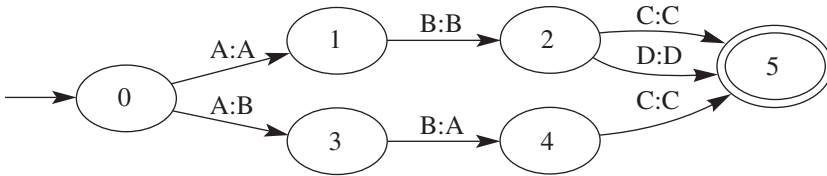


FIGURE 7. Finite-state transducer for morpheme order.

A second automaton, in Figure 8, generates the unattested pattern in which A and B are freely ordered, except when E occurs later in the word: {A-B-C-D, B-A-C-D, A-B-C-E} (*B-A-C-E). In this case, unlike in Fig. 7, the conditioner of the context-sensitive variation is not adjacent to the locus of variation (A and B), as C intervenes in all outputs.

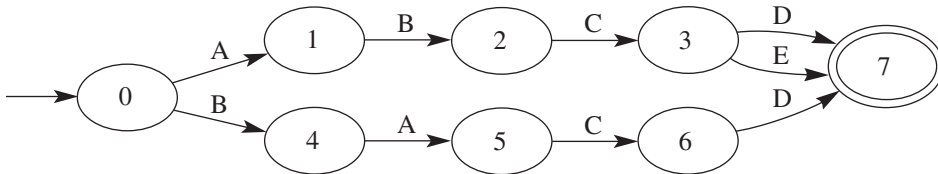


FIGURE 8. Automaton for an unattested pattern of variation.

A system of ranked or weighted bigrams can generate the first grammar (Fig. 7) but not the second (Fig. 8). To get variation between A-B-C-D and B-A-C-D, constraints A-B and B-A must be closely weighted (as in maxent grammar, e.g. Goldwater & Johnson 2003, or noisy harmonic grammar, e.g. Boersma & Pater 2008) or variably ranked (Anttila 1997a,b). There is no constraint that can force the A-B order in the context C-E but not C-D. Therefore, no grammar can rule out B-A-C-E while allowing A-B-C-D and B-A-C-D, as in Fig. 8.

This restrictiveness extends to output patterns for single inputs as well. For example, imagine a grammar in which input /A,B,C,D,E/ has exactly three outputs in free variation (A-B-C-D-E, B-A-D-C-E, B-D-A-E-C). Unlike the directed graph formalism in Figure 9, no bigram constraint grammar can generate those three candidates without substantially overgenerating other candidates (as confirmed by running in Wilson &

George 2008 the tableau with all 120 candidate permutations and all twenty pairwise-distinct bigram constraints). These examples are sufficient to show that interacting bigrams are not a brute-force approach like directed graphs.

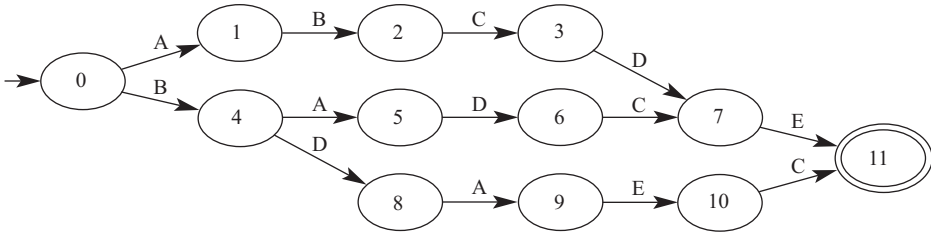


FIGURE 9. Automaton for an unattested pattern of variation for a single input.

Despite this formal restrictiveness of bigram morphotactics, there remain various typological tendencies in affix ordering that the theory does not address. To pick one, in a language with case and number affixes, number is almost always inside case (Greenberg 1963, Hawkins & Gilligan 1988). But under the present proposal, both CASE-NUMBER >> NUMBER-CASE and the opposite ranking are equally accessible to the learner. Even if we allowed that some faithfulness constraint was in the constraint set penalizing number outside of case (e.g. LINEARITY in Horwood 2002), as long as the learner is capable of positing a morphotactic constraint favoring the other order and candidates with the other order are generated, number outside of case is predicted to be synchronically accessible. This issue is a concern not only for my proposal, but also for any general theory addressing affix order variation and language-specific morphotactics, such as the theories mentioned in §7 and §8.

Nevertheless, it is often not clear to what extent such restrictions are imposed by the synchronic morphology as opposed to being diachronic artifacts with explanations in other domains (on piecing apart synchronic and diachronic explanation of typological generalizations, see e.g. Blevins 2004, Kiparsky 2006, Wilson 2006, Anderson 2008, and Moreton 2008). Is a language with number-outside-of-case morphology unlearnable? The existence of exceptions to the generalization (Konstanz Universals Archive, record 7; Plank & Filimonova 2000) demonstrates that this is not so. In that case, is there some kind of learning bias (Wilson 2006, Moreton 2008) disfavoring such a language? Or might its rarity be due entirely to the typical chronology of affixal fusion, whose causes might in turn be associated with other domains, such as the syntax (cf. Hawkins & Gilligan 1988, Bybee et al. 1990, Siewierska & Bakker 1996, Trommer 2003, etc.)? The Konstanz Universals Archive entry for the number/case generalization, for instance, includes the observation: ‘Number will always be grammaticalized before Case, hence bound Number exponents will always end up closer to the stem and bound Case exponents will always be more marginal’ (Frans Plank). Evaluating these possibilities is beyond the scope of this article, but if Plank’s suggestion is on the right track, hardwiring the case/number generalization into the synchronic morphology might not only be unnecessary and redundant (since the scarcity of the pattern would already be accounted for), but should in principle be empirically falsifiable (e.g. if learners exhibit no difficulty in acquiring the less common order).

In a similar vein, my proposal does not synchronically distinguish between the markedness of free variation vs. nonvariation, or transitivity vs. nontransitivity. Yet nonvariation and transitivity are impressionistically much better represented than their

counterparts typologically. In these cases, diachronic and/or functional (see e.g. Hay 2003, Hay & Plag 2004, and Plag & Baayen 2009) explanations of the asymmetries are especially plausible. I have argued in §6 that variation can evolve from a categorical system under special conditions (e.g. an infixing reduplicant reanalyzed as mobile due to morphotactic extension). But under more run-of-the-mill circumstances, the learner would typically have no motivation to innovate free variation (e.g. the Tagalog learners above did not learn variation for morphemes other than aspectual RED); the affixes would remain fixed in the order in which they were morphologized or the order determined by scope.

Likewise, if we assume that affixes are usually morphologized in a fixed chronological sequence (X, then Y, then Z), the combinations will be transitive by default (X-Y, X-Z, Y-Z, X-Y-Z), and the learner would typically have no reason to deviate from the transitive learning data. But under less frequent conditions, the learner might innovate nontransitivity (e.g. due to analogy, as in §6) or affixes might be morphologized in a nontransitive order. As an example of the latter, recall the case of nontransitivity in Lithuanian in 28 in which *si* ‘reflexive’ can either precede or follow the root, depending on whether a prefix (of a certain type) is present. According to Nevis and Joseph (1993), *si* was originally a second-position clitic, explaining its variable placement with respect to the root. If second-position clitics are uncommon, we would not expect this sort of situation to be morphologized often, even though it is entirely learnable when it is morphologized. In sum, any synchronic theory that cannot generate variation or nontransitivity is inadequate. But the explanation for the relative scarcity of variation and nontransitivity might reside in historical and functional considerations.

11. CONCLUSION. Bigram morphotactic constraints provide a constrained, sufficiently powerful, and demonstrably learnable means of implementing local morphological restrictions on the placement of RED in Tagalog and similar semantically unpredictable affix ordering restrictions in other languages. In fact, the model does better than covering the facts. When trained on an impoverished ‘core’ corpus on invariable (least marked output only) Tagalog data, the bigram learner correctly anticipates to a large extent which variant positions for RED should and should not be allowed, and in roughly which proportions. These simulations show that when a categorical training corpus exhibits certain characteristics, variation can be easier to learn than categoricity, motivating its diachronic emergence and synchronic stability.

The bigram schema is intended to be universal, but individual bigram constraints must be posited during learning (as in Pater 2007: ‘morpheme-specific constraints are constructed from universal constraints in the course of learning’). Roughly speaking, this entails that as the learner parses morphemes, he or she also pays attention to the local sequences in which those morphemes occur (see especially Poon et al. 2009 on morpheme parsing in a log-linear framework closely related to the maxent formalism in this article). Bigram morphotactics is intended to supplement, not replace, semantic factors in affix ordering such as scope. As discussed in §§1–2, scope plays a role in some (sub)systems, but not others. Because free variation is by definition not correlated with semantic effects, another theory, such as the present one, is necessary to account for its grammar.

REFERENCES

- ACKEMA, PETER, and AD NEELEMAN. 2004. *Beyond morphology: Interface conditions on word-formation*. Oxford: Oxford University Press.

- ACKEMA, PETER, and AD NEELEMAN. 2005. Word-formation in optimality theory. *Handbook of word-formation*, ed. by Pavol Štekauer and Rochelle Lieber, 285–313. Dordrecht: Springer.
- ALBÓ, XAVIER. 1964. *El Quechua a su Alcance*. La Paz: Alianza para el Progreso.
- ALSINA, ALEX. 1999. Where's the mirror principle? *The Linguistic Review* 16.1–42.
- ANDERSON, STEPHEN. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4.1–32.
- ANDERSON, STEPHEN. 2008. The logical structure of linguistic theory. *Language* 84.795–814.
- ANTTILA, ARTO. 1997a. Deriving variation from grammar. *Variation, change and phonological theory*, ed. by Frans Hinskens, Roeland van Hout, and Leo Wetzels, 35–68. Amsterdam: John Benjamins.
- ANTTILA, ARTO. 1997b. *Variation in Finnish phonology and morphology*. Stanford, CA: Stanford University dissertation.
- ARONOFF, MARK, and ZHENG XU. 2010. A realization optimality-theoretic approach to affix order. *Morphology*, to appear. Online: roa.rutgers.edu/view.php3?id=1518.
- BAKER, MARK. 1985. The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16.373–415.
- BARONI, MARCO. 2003. Distribution-driven morpheme discovery: A computational/experimental study. *Yearbook of Morphology* 2003.213–48.
- BECK, D. 2007. Variable ordering of affixes in Upper Necaxa Totonac. Paper presented at the 12th Workshop on Structure and Constituency of the Languages of the Americas, University of Lethbridge.
- BECKER, MICHAEL; JOE PATER; and CHRISTOPHER POTTS. 2007. OT-Help 1.2. Software package. Amherst: University of Massachusetts, Amherst.
- BICKEL, BALTHASAR; GOMA BANIADE; MARTIN GAENZLE; ELENA LIEVEN; NETRA PRASAD PAUDYAL; ICHCHHA PURNA RAI; MANOJ RAI; NOVEL KISHORE RAI; and SABINE STOLL. 2007. Free prefix ordering in Chintang. *Language* 83.43–73.
- BLEVINS, JULIETTE. 2001. *Nhanda: An aboriginal language of Western Australia*. Honolulu: University of Hawai'i Press.
- BLEVINS, JULIETTE. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.
- BOBALJIK, JONATHAN D. 2000. Adverbs: The hierarchy paradox. *Glott International* 4.9–10.27–28.
- BOERSMA, PAUL. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam (IFA)* 21.43–58.
- BOERSMA, PAUL, and BRUCE HAYES. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.45–86.
- BOERSMA, PAUL, and JOE PATER. 2008. Convergence properties of a gradual learning algorithm for harmonic grammar. Amsterdam: University of Amsterdam, and Amherst: University of Massachusetts, Amherst, MS.
- BOOIJ, GEERT, and ROCHELLE LIEBER. 1993. On the simultaneity of morphological and prosodic structure. *Studies in lexical phonology* (Phonetics and phonology 4), ed. by Sharon Hargus and Ellen Kaisse, 23–44. San Diego: Academic Press.
- BUELL, LESTON; MARIAME SY; and HAROLD TORRENCE. 2008. The syntax of mirror principle violations in Wolof. Handout from the 2008 International Morphology Meeting, Vienna, February 2008.
- BYBEE, JOAN L. 1985. *Morphology: A study of the relation between meaning and form*. (Typological studies in language 9.) Amsterdam: John Benjamins.
- BYBEE, JOAN L.; WILLIAM PAGLIUCA; and REVERE D. PERKINS. 1990. On the asymmetries in the affixation of grammatical material. *Studies in typology and diachrony*, ed. by William Croft, Keith Denning, and Suzanne Kemmer, 1–42. Amsterdam: John Benjamins.
- CABALLERO, GABRIELA. 2008. *Choguita Rarámuri (Tarahumara) phonology and morphology*. Berkeley: University of California, Berkeley dissertation.
- CABALLERO, GABRIELA. 2010. Multiple exponence and the phonology-morphology interface. *North East Linguistic Society (NELS)* 39, to appear.
- CABALLERO, GABRIELA. 2011. Scope, phonology and templates in an agglutinating language: Choguita Rarámuri (Tarahumara) variable suffix ordering. *Morphology*, to appear.

- CARRIER, JILL LOUISE. 1979. *The interaction of morphological and phonological rules in Tagalog: A study in the relationship between rule components in grammar*. Cambridge, MA: MIT dissertation.
- CHOMSKY, NOAM. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- CHOMSKY, NOAM. 2001. Derivation by phase. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- COLE, JENNIFER. 1994. *A prosodic theory of reduplication*. Stanford, CA: Stanford University dissertation.
- CONDORAVDI, CLEO, and PAUL KIPARSKY. 1998. Optimal order and scope. Paper presented at The Lexicon in Focus, Heinrich-Heine Universität Düsseldorf, August 1998.
- COURTNEY, ELLEN H., and MURIEL SAVILLE-TROIKE. 2002. Learning to construct verbs in Navajo and Quechua. *Journal of Child Language* 29.623–54.
- DOWNING, LAURA. 1998. Prosodic misalignment and reduplication. *Yearbook of Morphology* 1997.83–120.
- ELMAN, JEFFREY. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48.71–99.
- EMBICK, DAVID. 2007. Linearization and local dislocation: Derivational mechanics and interactions. *Linguistic Analysis* 33.303–36.
- EMBICK, DAVID. 2008. Localism versus globalism in morphology and phonology. Philadelphia: University of Pennsylvania, MS.
- EMBICK, DAVID, and ROLF NOYER. 2001. Movement operations after syntax. *Linguistic Inquiry* 32.555–95.
- FABB, NIGEL. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6.527–39.
- FRENCH, KOLEEN. 1988. *Insights into Tagalog: Reduplication, infixation and stress from nonlinear phonology*. Dallas: SIL International and the University of Texas, Arlington.
- GANSNER, EMDEN, and STEPHEN NORTH. 1999. An open graph visualization system and its applications to software engineering. *Software—Practice and Experience* 30.1203–33.
- GOLDSMITH, JOHN. 2001. Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27.153–98.
- GOLDWATER, SHARON, and MARK JOHNSON. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–20. Stockholm: Stockholm University Department of Linguistics.
- GOOD, JEFF. 2003. *Strong linearity: Three case studies towards a theory of morphosyntactic templatic constructions*. Berkeley: University of California, Berkeley dissertation.
- GOOD, JEFF. 2005. Reconstructing morpheme order in Bantu: The case of causativization and applicativization. *Diachronica* 22.3–57.
- GOOD, JEFF. 2007. Strong linearity, weak linearity, and the typology of templates. *New challenges in typology: Broadening the horizons and redefining the foundations*, ed. by Matti Miestamo and Bernhard Wälchli, 11–33. Berlin: Mouton de Gruyter.
- GOOD, JEFF, and ALAN YU. 1998. Suffix ordering and variability in Turkish. Paper presented at the 25th annual meeting of the Berkeley Linguistics Society, Berkeley, California.
- GOOD, JEFF, and ALAN YU. 2005. Morphosyntax of two Turkish subject pronominal paradigms. *Clitic and affix combinations*, ed. by Lorie Heggie and Francisco Ordóñez, 315–42. Amsterdam: John Benjamins.
- GREENBERG, JOSEPH. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of human language*, ed. by Joseph Greenberg, 73–113. Cambridge, MA: MIT Press.
- GRIMES, JOSEPH E. 1983. *Affix positions and co-occurrence: The PARADIGM program*. Dallas: Summer Institute of Linguistics.
- HALLE, MORRIS. 1990. An approach to morphology. *North East Linguistic Society (NELS)* 20.150–84.
- HALLE, MORRIS, and ALEC MARANTZ. 1993. Distributed morphology and the pieces of inflection. *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 111–76. Cambridge, MA: MIT Press.
- HARGUS, SHARON, and SIRI G. TUTTLE. 1997. Augmentation as affixation in Athabaskan languages. *Phonology* 14.177–220.

- HAWKINS, JOHN, and GARY GILLIGAN. 1988. Prefixing and suffixing universals in relation to basic word order. *Lingua* 74.219–59.
- HAY, JENNIFER. 2003. *Causes and consequences of word structure*. London: Routledge.
- HAY, JENNIFER, and INGO PLAG. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22.565–96.
- HAYES, BRUCE, and COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.
- HERRERO, JOAQUIN, and FREDERICO SÁNCHEZ DE LOZADA. 1978. *Gramática Quechua: Estructura del quechua boliviano contemporáneo*. Cochabamba: Editorial Universo.
- HORWOOD, GRAHAM. 2002. Precedence faithfulness governs morpheme position. *West Coast Conference on Formal Linguistics (WCCFL)* 21.166–79.
- HYMAN, LARRY. 1994. Conceptual issues in the comparative study of the Bantu verb stem. *Topics in African linguistics*, ed. by Salikoko Mufwene and Lioba Moshi, 3–34. Amsterdam: John Benjamins.
- HYMAN, LARRY. 2003. Suffix ordering in Bantu: A morphocentric approach. *Yearbook of Morphology* 2002.245–81.
- INKELAS, SHARON. 1993. Nimboran position class morphology. *Natural Language and Linguistic Theory* 11.559–624.
- INKELAS, SHARON. 2000. Infixation obviates backcopying in Tagalog. Paper presented at the annual meeting of the Linguistic Society of America, Chicago, IL, January 2000.
- INKELAS, SHARON, and GABRIELA CABALLERO. 2008. An inside-out approach to multiple exponence. Paper presented at the Berkeley Workshop on Affix Ordering, October 2008.
- INKELAS, SHARON, and CHERYL ZOLL. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.
- JÄGER, GERHARD. 2007. Maximum entropy models and stochastic optimality theory. *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen, 467–79. Stanford, CA: CSLI Publications.
- JAKER, ALESSANDRO. 2006. Split subject agreement and morphological typology. Stanford, CA: Stanford University, MS.
- JOHNSON, MARK. 2002. Optimality-theoretic lexical functional grammar. *The lexical basis of sentence processing: Formal, computational, and experimental issues*, ed. by Paula Merlo and Susan Stevenson, 59–74. Amsterdam: John Benjamins.
- JOSEPH, BRIAN. 2007. What Greek can tell us about the remaking of verb endings. Columbus: The Ohio State University, MS. Online: www.ling.ohio-state.edu/~bjoseph/publicat.htm.
- JURAFSKY, DANIEL, and JAMES MARTIN. 2000. *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.
- KIPARSKY, PAUL. 1982. From cyclic to lexical phonology. *The structure of phonological representations*, part 1, ed. by Harry van der Hulst and Norval Smith, 131–75. Dordrecht: Foris.
- KIPARSKY, PAUL. 2000. Opacity and cyclicity. *The Linguistic Review* 17.351–67.
- KIPARSKY, PAUL. 2006. Amphichronic linguistics vs. evolutionary phonology. *Theoretical Linguistics* 32.217–36.
- KROEGER, PAUL. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications.
- KURISU, KAZUTAKA. 2001. *The phonology of morpheme realization*. Santa Cruz: University of California, Santa Cruz dissertation.
- LUUTONEN, JORMA. 1997. *The variation of morpheme order in Mari declension*. Helsinki: Suomalais-Ugrilainen Seura.
- MACBRIDE, ALEX. 2004. *A constraint-based approach to morphology*. Los Angeles: University of California, Los Angeles dissertation.
- MCCARTHY, JOHN J. 2011. Pausal phonology and morpheme realization. *Prosody matters: Essays in honor of Lisa Selkirk*, ed. by Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara. London: Equinox, to appear.
- MCCARTHY, JOHN J., and ALAN S. PRINCE. 1994. Generalized alignment. *Yearbook of Morphology* 1993.79–153.

- MCCARTHY, JOHN J., and ALAN S. PRINCE. 1995. Faithfulness and reduplicative identity. *Papers in optimality theory* (University of Massachusetts occasional papers in linguistics 18), ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249–384. Amherst: University of Massachusetts Graduate Linguistics Student Association.
- McFARLAND, TERESA. 2005. The inflectional system of Totonaco de Filomeno Mata. Paper presented at the annual meeting of the Linguistic Society of America, Oakland, CA, January 2005.
- McFARLAND, TERESA. 2007. Free affix order in Filomeno Mata Totonac. Paper presented at the annual meeting of the Linguistic Society of America, Anaheim, CA, January 2007.
- MCPHERSON, LAURA, and MARY PASTER. 2009. Evidence for the mirror principle and morphological templates in Luganda affix ordering. *Proceedings of the 39th annual Conference on African Linguistics*, ed. by Akinloye Ojo and Lioba Moshi, 56–66. Somerville, MA: Cascadilla.
- MERCADO, RALPH. 2007. The mystery of the *mag-* prefix: Phases and reduplication in Tagalog. Paper presented at the 14th meeting of the Austronesian Formal Linguistics Association, Montreal, Quebec.
- MOHANAN, K. P. 1986. *The theory of lexical phonology*. Dordrecht: Reidel.
- MORETON, ELLIOTT. 1999. Non-computable functions in optimality theory. Amherst: University of Massachusetts, Amherst, MS.
- MORETON, ELLIOTT. 2008. Analytic bias and phonological typology. *Phonology* 25.83–127.
- MUNRO, PAMELA, and JASON RIGGLE. 2004. Productivity and lexicalization in Pima compounds. *Berkeley Linguistics Society* 30.114–26.
- MUYSKEN, PIETER. 1981a. Quechua causatives and logical form: A case study in markedness. *Theory of markedness in generative grammar*, ed. by Adriana Belletti, Luciana Brandi, and Luigi Rizzi, 445–73. Pisa: Scuola Normale Superiore di Pisa.
- MUYSKEN, PIETER. 1981b. Quechua word structure. *Binding and filtering*, ed. by Frank Heny, 279–327. Cambridge, MA: MIT Press.
- MUYSKEN, PIETER. 1986. Approaches to affix order. *Linguistics* 24.629–43.
- MUYSKEN, PIETER. 1988. Affix order and interpretation: Quechua. *Morphology and modularity*, ed. by Martin Everaert, Arnold Evers, Riny Huybregts, and Mieke Trommelen, 259–79. Dordrecht: Foris.
- NAGY, NAOMI, and BILL REYNOLDS. 1997. Optimality theory and word-final deletion in Faetar. *Language Variation and Change* 9.37–55.
- NEVIS, JOEL, and BRIAN JOSEPH. 1993. Wackernagel affixes: Evidence from Balto-Slavic. *Yearbook of Morphology* 1992.93–111.
- NOYER, ROLF. 1993. Optimal words: Towards a declarative theory of word formation. Princeton, NJ: Princeton University, MS.
- PASTER, MARY. 2006a. *Phonological conditions on affixation*. Berkeley: University of California, Berkeley dissertation.
- PASTER, MARY. 2006b. Pulaar verbal extensions and phonologically driven affix order. *Yearbook of Morphology* 2005.155–99.
- PATER, JOE. 2007. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. *Phonological argumentation: Essays on evidence and motivation*, ed. by Steve Parker, 123–54. London: Equinox.
- PATER, JOE. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33.999–1035.
- PESETSKY, DAVID A. 1985. Morphology and logical form. *Linguistic Inquiry* 16.193–246.
- PETERSON, DAVID A. 1994. Multiple exponence and morphosyntactic redundancy. *West Coast Conference on Formal Linguistics (WCCFL)* 12.83–100.
- PLAG, INGO, and HARALD BAAYEN. 2009. Suffix ordering and morphological processing. *Language* 85.109–52.
- PLANK, FRANS, and ELENA FILIMONOVA. 2000. The Universals Archive: A brief introduction to prospective users. *Sprachtypologie und Universalienforschung* 53.109–23.
- POON, HOIFUNG; COLIN CHERRY; and KRISTINA TOUTANOVA. 2009. Unsupervised morphological segmentation with log-linear models. *Human Language Technologies: The 2009 annual conference of the North American Chapter of the ACL*, 209–17. Boulder, CO: Association for Computational Linguistics.

- POTTS, CHRISTOPHER; JOE PATER; KAREN JESNEY; RAJESH BHATT; and MICHAEL BECKER. 2010. Harmonic grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27.77–117.
- PRESS, WILLIAM; BRIAN FLANNERY; SAUL TEUKOLSKY; and WILLIAM VETTERLING. 1992. *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press.
- PRINCE, ALAN, and PAUL SMOLENSKY. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- RACKOWSKI, ANDREA. 1999. Morphological optionality in Tagalog aspectual reduplication. *MIT Working Papers in Linguistics (Papers on morphology and syntax, cycle two)* 34.107–36.
- RAMOS, TERESITA V., and MARIA LOURDES S. BAUTISTA. 1986. *Handbook of Tagalog verbs: Inflections, modes, and aspects*. Honolulu: University of Hawai'i Press.
- REYNOLDS, WILLIAM. 1994. *Variation and phonological theory*. Philadelphia: University of Pennsylvania dissertation.
- RICE, KEREN. 1993. The structure of the Slave (Northern Athapaskan) verb. *Studies in lexical phonology* (Phonetics and phonology 4), ed. by Sharon Hargus and Ellen Kaisse, 145–71. San Diego: Academic Press.
- RICE, KEREN. 2000. *Morpheme order and semantic scope: Word formation in the Athabaskan verb*. Cambridge: Cambridge University Press.
- RIGGLE, JASON, and COLIN WILSON. 2004. Local optionality. *North East Linguistic Society (NELS)* 35.539–50.
- ROSE, SHARON. 1997. *Theoretical issues in comparative Ethio-Semitic phonology and morphology*. Montreal: McGill University dissertation.
- ROSENBLATT, FRANK. 1958. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65.386–408.
- SABIMANA, FIRMARD. 1986. *The relational structure of the Kirundi verb*. Bloomington: Indiana University dissertation.
- SAMEK-LODOVICI, VIERI. 1993. Morphological gemination. Paper presented at ROW-1, Rutgers University, October 1993.
- SAMUELS, BRIDGET. 2006. Reduplication and verbal morphology in Tagalog. Paper presented at the Harvard Linguistics Theory Reading Group, October 2006.
- SCHACHTER, PAUL, and FE OTANES. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
- SENN, ALFRED. 1966. *Handbuch der litauischen Sprache*. Heidelberg: Carl Winters Universitätsverlag.
- SIEWIERSKA, ANNA, and DIK BAKKER. 1996. The distribution of subject and object agreement and word order type. *Studies in Language* 20.115–61.
- SKINNER, TOBIN R. 2008. Morphological optionality in reduplication: A lowering account. *West Coast Conference on Formal Linguistics (WCCFL)* 26.420–28.
- SMEETS, INEKE. 1989. *A Mapuche grammar*. Leiden: University of Leiden dissertation.
- SMOLENSKY, PAUL, and GÉRALDINE LEGENDRE. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.
- SPENCER, ANDREW. 1991. *Morphological theory*. Oxford: Blackwell.
- STAIRS, EMILY F., and BARBARA E. HOLLENBACH. 1981. Gramática huave. *Diccionario huave de San Mateo del Mar*, ed. by G. A. S. Kreger and Emily F. Stairs, 283–391. Mexico City: SIL International.
- STEVENS, ALAN M. 1971. Fixed morpheme order. *Linguistic Inquiry* 2.420–21.
- STUMP, GREGORY. 2006. Template morphology. *Encyclopedia of language and linguistics*, vol. 12, ed. by Keith Brown, 559–63. Oxford: Elsevier.
- TRAVIS, LISA DEMENA. 1996. The syntax of achievements. Montreal: McGill University, MS.
- TRAVIS, LISA DEMENA. 2007. *Inner aspect: Articulation of the VP*. Dordrecht: Kluwer.
- TROMMER, JOCHEN. 2003. The interaction of morphology and syntax in affix order. *Yearbook of Morphology* 2002.283–324.
- USSISHKIN, ADAM, and ANDREW WEDEL. 2002. Neighborhood density and the root-affix distinction. *North East Linguistic Society (NELS)* 32.539–49.
- VAN DE KERKE, SIMON. 1996. *Affix order and interpretation in Bolivian Quechua*. Meppel: Krips Repro.

- VAUX, BERT. 2002. Iterativity and optionality. Paper presented at the 33rd conference of the North East Linguistic Society (NELS), November 2002.
- WALKER, RACHEL. 1998. *Nasalization, neutral segments, and opacity effects*. Santa Cruz: University of California, Santa Cruz dissertation. [Published, New York: Garland, 2000.]
- WATTERS, DAVID E. 2006. Notes on Kusunda grammar: A language isolate of Nepal. *Himalayan Linguistics Archive* 3.1–182.
- WILSON, COLIN. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–82.
- WILSON, COLIN, and BENJAMIN GEORGE. 2008. Maxent grammar tool. Software package. Los Angeles: University of California, Los Angeles.
- WOLF, MATTHEW. 2008. *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. Amherst: University of Massachusetts, Amherst dissertation.
- ZURAW, KIE. 1996. Floating phonotactics: Infixation and reduplication in Tagalog loanwords. Los Angeles: University of California, Los Angeles masters thesis.
- ZURAW, KIE. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28.417–72.

UCLA Department of Linguistics
3125 Campbell Hall
Los Angeles, CA 90095-1543
[kmryan@humnet.ucla.edu]

[Received 16 July 2009;
accepted 1 January 2010]