

Intersecting constraint families: an argument for Harmonic Grammar

Kie Zuraw and Bruce Hayes

UCLA

Kie Zuraw:

kie@ucla.edu

Bruce Hayes:

bhayes@humnet.ucla.edu

both:

Department of Linguistics

3125 Campbell Hall

UCLA

Los Angeles, CA 90095-1543

USA

Intersecting constraint families: an argument for Harmonic Grammar

In the analysis of free variation in phonology, we often encounter the effects of INTERSECTING CONSTRAINT FAMILIES: there are two independent families of constraints, each of which has a quantifiable effect on the outcome. A challenge for theories is to account for the patterns that emerge from such intersection. We address three cases: Tagalog nasal substitution, French liaison/élision, and Hungarian vowel harmony, using corpus data. We characterize the patterns we find as across-the-board effects in both dimensions, restrained by floor and ceiling limits. We analyze these patterns using several formal frameworks, and find that an accurate account is best based on Harmonic Grammar (in one of its two primary quantitative implementations). Our work also suggests that that certain lexical distinctions treated as discrete by classical phonological theory (e.g. ‘*h aspiré*’ vs. ordinary vowel-initial words of French) are in fact gradient and require quantitative treatment.*

* We would like to thank Andy Lin of UCLA’s statistical consulting group for help without which the project would not have been feasible. Giorgio Magri kindly formulated and proved the theorem we use in §3.8.1. We also thank for their helpful advice Arto Anttila and Paul Smolensky, as well as audiences at the West Coast Conference on Formal Linguistics (Santa Cruz), the Workshop on Altaic Formal Linguistics (Cornell), the UCLA-USC phonology seminar, Seoul National University, and Yale University. The authors take full responsibility for remaining errors.

1. INTRODUCTION. A key debate in constraint-based linguistic frameworks concerns RANKING versus WEIGHTING. Optimality Theory (Prince & Smolensky 1993/2004) uses strict ranking: Candidate A is preferred to Candidate B if the highest-ranked constraint that distinguishes between them prefers A. In Harmonic Grammar (Legendre, Miyata & Smolensky 1990, Legendre, Sorace & Smolensky 2006, Potts et al. 2010, and others), the constraints bear numeric weights, and the winner is the candidate with the best Harmony value (a weighted sum of constraint violations).

In both approaches there are ways, discussed below, of elaborating the core ideas to assign probability distributions to outputs, rather than predicting a single winner. Such grammars can be used to model free variation and gradient intuitions, increasing the realism of analysis and its engagement with data. Some of the probabilistic frameworks are affiliated with learning algorithms, which make it possible to address more directly the core theoretical issue (e.g. Chomsky 1965) of how language is acquired. However, grammars constructed within these frameworks have similar behavior, capturing both extreme and intermediate rates of variation. This makes it hard to produce empirical arguments in favor of one particular framework. Until now, researchers have tended to adopt one framework or another on the grounds of computational ease, or the convergence properties of the available learning algorithms (e.g. Pater 2008, Magri 2012). We address this difficulty here by examining how the frameworks fare on INTERSECTING CONSTRAINT FAMILIES.

We argue that the quantitative patterns we observe in three case studies have a natural account with constraint weighting, but not with constraint ranking. Results are similar for two implementations of Harmonic Grammar, Noisy HG and MaxEnt. We also consider and reject models in which the constraints from the two dimensions do not interact directly but instead are placed in different components (decision-tree models).

A secondary point we will make concerns LEXICAL PROPENSITY, the tendency of individual lexical items toward one phonological behavior. We will see this particularly in the case study below of French, where different words have different, idiosyncratic tendencies to require a preceding syllable boundary, and in the case study of Tagalog, where different prefixes have different tendencies to resist a phonological rule. We will show that lexical propensity is well modeled with lexically specific constraint weights.

1.1 INTERSECTING CONSTRAINT FAMILIES. Intersecting constraint families arise when a phonological outcome is influenced by two independent dimensions, with each dimension referred to by a different set (FAMILY) of constraints. For instance, we will see below that the frequency of front or back suffix allomorphs in Hungarian vowel harmony is determined by one set of constraints that refer to the vowels of the stem, and a second that refer to the stem-final consonant(s). The two constraint sets are orthogonal because there is no necessary connection in Hungarian between what vowels a stem may have and what consonants it ends in. It is an empirical question how the vowel-constraint family and the consonant-constraint family interact in determining harmony.

The question of intersecting constraint families can be visualized as a rectangular array in which the rows and columns represent constraints of the two families, as schematized in Figure 1, where there are constraints applicable to different consonant places, and constraints applicable to different vowel heights. Each data cell (shown in dark grey) occurs at the intersection of two constraints, and holds the rate of the phonological outcome for words that those two constraints apply to. In this hypothetical illustration, the outcome is vowel devoicing, and it depends on both consonant place and vowel height.¹

Figure 1 <INSERT Figure 1 HERE>

On standard scientific grounds, we would hardly want to set up a grammar that stipulates the outcome rate for each cell separately, with nine separate parameters. Rather, we would prefer a system that assigns the right properties (rankings or weights) to each row and column, and lets the behavior of the individual cells follow from the general theory of constraint interaction.

As we will see, the available theories make quite different predictions about how intersecting constraint families should behave. It will emerge that Harmonic Grammar, in two different versions (MaxEnt grammars and Noisy Harmonic Grammar) makes predictions that match well with data, and the alternative theories are considerably less successful.

1.2 CASES EXAMINED. To make our inquiry possible, we require phenomena with the following four properties. First, there must be variable output that depends predictably on some factors. Second, the factors must belong to two (or more) orthogonal families of constraints. Third, the degree of variation should be substantial in both families, ideally reaching floor and ceiling levels (0 and 1) for both. Lastly, there must be ample corpus data: if the intersecting constraint families have n and m members, respectively, the corpus must be large enough to populate each cell of an n by m rate matrix with a substantial number of observations.

We will present case studies from Tagalog, French, and Hungarian that meet or approximate the above criteria. We will see that Harmonic Grammar is successful regardless of whether the intersecting constraint families are antagonistic (one dimension consists of constraints favoring one candidate, and the other dimension of constraints favoring the other candidate), synergistic (both dimensions consist of constraints favoring the same candidate, against a single opposing constraint that applies in all cases), or a mix. The French case involves synergistic constraint families, and Hungarian and Tagalog both involve a mix.

2. TAGALOG NASAL SUBSTITUTION

2.1 NASAL SUBSTITUTION. The process known as nasal substitution is widespread in Western Austronesian languages (Newman 1984, Blust 2004, Pater 2001). In the Tagalog version, analyzed in detail in Zuraw 2010, the final /ŋ/ of a prefix fuses with the initial obstruent of a stem to form a nasal that has the original obstruent's place of articulation.

(1) Tagalog nasal substitution

/maŋ-bigáj/	→ [ma-migáj]	'to distribute'
/maŋ-súlat/	→ [mà-nulát]	'to write professionally'
/ma-paŋ-kamkám/	→ [ma-pa-ŋamkám]	'rapacious'

Nasal substitution can apply to /p/ and /b/, yielding [m], to /t, d, s/, yielding [n], and to [k, g, ʔ], yielding [ŋ]. When nasal substitution does not apply, the prefix /ŋ/ generally assimilates in place to the following obstruent.

(2) Tagalog nasal substitution: examples of non-application

/paŋ-poʔók/	→ [pam-poʔók]	‘local’
/paŋ-súlat/	→ [pan-súlat]	‘writing instrument’
/maŋ-RED-kúlam/	→ [maŋ-ku-kúlam]	‘witch’

Although every Tagalog obstruent has the possibility of undergoing nasal substitution or not, there are two strong phonological trends: voiceless obstruents undergo more than voiced, and, especially within the voiced obstruents, fronter (labial) consonants undergo at higher rates and backer (velar) at lower rates. The mosaic plot in Figure 2, taken from Zuraw 2010, shows dictionary data (English 1986). For each place/voicing combination, the number of words listed in the dictionary as substituted or unsubstituted is plotted; words listed with two pronunciations are omitted in this plot. For example, the dictionary lists 253 words whose stem begins with /p/ that undergo nasal substitution, and 10 whose stem begins with /p/ that do not undergo. The width of each column is proportional to the number of words making it up. Data from a corpus of mixed web text (blogs, discussion forums, newspapers; Zuraw 2009) show similar trends; see Zuraw 2010.

Figure 2<INSERT Figure 2 HERE>

2.2 NASAL SUBSTITUTION RATES WITH DIFFERENT PREFIXES. We have now seen the first of our constraint families, phonological constraints that concern consonant place and voicing. The second family will concern morphology. Figure 2 included all relevant words in the dictionary, regardless of morphology. If we split up the data by morphology, however, we find that the prefix constructions differ greatly in the frequency with which they trigger nasal substitution. Figure 3 is a mosaic plot, based on Zuraw 2010,² that shows nasal substitution rates for the six most type-frequent prefixing constructions in the dictionary, and includes words listed in the dictionary as having both substituted and unsubstituted pronunciations (in grey).

Figure 3<INSERT Figure 3 HERE>

2.3 INTERACTION OF THE TWO FACTORS. Since both consonant identity and prefix identity strongly influence the frequency of substitution, we can examine how they interact, with the ultimate goal of modeling this interaction in various frameworks. The interaction data are given in Figure 4, where stem-initial consonant is plotted on the horizontal axis and each trace represents one prefix construction. If a word is listed in the dictionary as variable, it contributes 0.5 to the non-substituting count for the relevant consonant and prefix, and 0.5 to the substituting count.

Figure 4<INSERT Figure 4 HERE>

There are two essential observations to be made. First, by and large, both the voicing and place effects hold **ACROSS THE BOARD**: they hold true not just of the data as a whole, but to a fair extent for each individual prefix. This can be seen in the roughly parallel course of the lines plotted in Figure 4. The prefix effects are also, to a fair degree, across-the-board: they hold for most of the consonants taken individually, which can be seen in the dearth of crossing lines. Thus, particular combinations of consonant and prefix do not overturn the general patterns.

Our second point is that where across-the-board patterning is not complete, this is because it is limited by **FLOOR** and **CEILING** effects. For instance, the prefix with the highest rate (*maŋ-*) flattens out place differences among *p*, *t/s*, *k* against the ceiling of 100%, and the prefix with the lowest rate (*paŋ-* reservational) flattens out most of the place difference among *b*, *d*, *g* against the floor. The intermediate prefixes show greater consonantal distinctions. By the same token, morphological differences are greatest for consonants with intermediate rates of nasal substitution (e.g. *b*), whereas the consonant with the highest rate (*p*) compresses four of the prefixes against the ceiling, and the consonant with the lowest rate (*g*) compresses all of the prefixes nearly against the floor.

The basic pattern—across-the-board effects in both dimensions, restrained by floor and ceiling effects—is found, we claim, in all of the data we examine in this article. Our goal is to study how grammar models can capture such effects. To start, we must set up a constraint system that permits the models to engage with the Tagalog data.

2.4 CONSTRAINTS FOR GRAMMAR MODELING. The Tagalog models below all use the same constraint set, adapted from Zuraw 2010. The basic setup is a standard one in Optimality Theory: markedness constraints that favor a particular change are opposed by faithfulness constraints. We set up the following markedness constraint as the fundamental force behind nasal substitution.

(3) **NASSUB:** Assess one violation for any nasal+obstruent sequence, where + is a morpheme boundary within a word.

When the stem begins with a voiceless obstruent, an additional markedness constraint further favors substitution.

(4) ***NC_o:** Assess one violation for every sequence of a nasal and a voiceless obstruent.

This constraint is responsible for higher rates of nasal substitution seen in *p/t/s/k*. See Zuraw 2010 for its phonetic motivation, Pater 1999 and Pater 2001 for its role in nasal substitution cross-linguistically.

The opposing faithfulness constraints, following Pater (1999), are all UNIFORMITY constraints (McCarthy & Prince 1995) that ban merging two input segments into one output segment, hence penalizing nasal substitution. Our UNIFORMITY constraints are indexed to each prefix construction, to capture the prefix differences observed above.

(5) Faithfulness constraints penalizing nasal substitution

- a. UNIF-*maŋ*-OTHER: One segment from input *maŋ*-*other* and a distinct input segment must not correspond to the same output segment
- b. UNIF-*paŋ*-RED-: similarly for *paŋ*-RED-
- c. UNIF-*maŋ*-RED-: similarly for *maŋ*-RED-
- d. UNIF-*maŋ*-ADV: similarly for *maŋ*-*adv*
- e. UNIF-*paŋ*-NOUN: similarly for *paŋ*-*noun*
- f. UNIF-*paŋ*-RES: similarly for *paŋ*-*res*

We capture the place-of-articulation differences with markedness constraints that militate against nasal substitution by penalizing roots that begin with the nasal segments that result.

- (6) Markedness constraints penalizing nasal substitution
- a. $*_{[\text{root } m/n/\eta]}$: one violation for every root that begins with a nasal
 - b. $*_{[\text{root } n/\eta]}$: one violation for every root that begins with [n] or [ŋ] (penalizes substitution on coronals and velars)
 - c. $*_{[\text{root } \eta]}$: one violation for every root that begins with [ŋ] (penalizes substitution on velars)

These constraints express a general tendency for Tagalog roots not to begin with nasals: only 5% of disyllabic native roots in English's (1986) dictionary begin with a nasal, while 22% of root-medial, and 16% of root-final consonants are nasal. This trend is stronger for backer nasals. We regard the $*_{[\text{root } \text{NASAL}]}$ constraints as special cases of general constraints against domain-initial sonorants (e.g. Prince & Smolensky's [1993/2004] family of MARGIN/ λ constraints; see Flack 2007 on domain-initial [ŋ] in particular). Their formulation assumes that the nasal resulting from nasal substitution is root-initial; for support for this assumption from reduplication see Zuraw 2010.

The consonant-sensitive constraints in 4 and 6 can be thought of as one constraint family, and the prefix-sensitive constraints in 5 as another.

We provide sample tableaux used in our simulations below in 7, showing the input, both output candidates, and all constraint violations. For the full set of 36 tableaux, see the OTSoft input files provided as supplemental materials.

- (7) Sample Tagalog tableaux
- <INSERT Tableau 7 HERE>

With these constraints in place, we can now turn to a variety of models and see how they perform in accounting for the data pattern.

2.5 ANALYSIS IN HARMONIC GRAMMAR. Harmonic Grammar is rooted in early work in general cognition by Smolensky (1986); the first application to language is Legendre, Miyata & Smolensky 1990. The key idea behind all forms of Harmonic Grammar is that constraints are not RANKED but WEIGHTED; that is, each constraint bears a real number that reflects its importance in the selection of winning candidates.

All forms of Harmonic Grammar start with the weighted sum of the constraint violations of each candidate, often called the HARMONY of the candidate. The two flavors of Harmonic Grammar we will cover here, Maximum Entropy and Noisy Harmonic Grammar, differ in how the harmony values are cashed out as quantitative predictions about frequency.

MAXIMUM ENTROPY MODEL. The roots of MaxEnt can be seen a variety of fields: physics and information theory (Jaynes 1957), cognitive science (Smolensky 1986), and computer science (Berger, Della Pietra & Della Pietra 1996, Della Pietra, Della Pietra & Lafferty 1997). The implementation of basic MaxEnt ideas within the GEN + EVAL architecture of Optimality Theory (Prince & Smolensky 1993/2004) first appeared in Goldwater and Johnson 2003.³ MaxEnt renders Harmonic Grammar probabilistic by means of a formula that converts harmony values to output probability; this is given in 8 (Della Pietra, Della Pietra & Lafferty 1997: 1) and unpacked immediately below.⁴

$$(8) \quad \text{Formula for probability in MaxEnt: } p(\omega) = \frac{1}{Z} e^{-\sum_i w_i C_i(\omega)} \quad \text{where } Z = \sum_j e^{-\sum_i w_i C_i(\omega_j)}$$

We illustrate how this works with an annotated example tableau in 9. When a /t/-initial stem is prefixed with noun-forming *paŋ-*, the nasal-substituted output candidate wins for about two-thirds of words. An example is /paŋ+tańím/ ‘something used for planting’. This particular word can surface as either [pananím] or [pantaním], but the 67% figure represents the behavior of the language as a whole: in the dictionary data there are 60.5 words with substitution (counting free variants by halves) and 29.5 without. Instead of asterisks, each cell shows a candidate’s number of violations.

(9) Sample MaxEnt tableau⁵

<INSERT Tableau 9 HERE>

The purpose of the grammar is to generate the probabilities in the rightmost column of 9. The first step is to compute each candidate's harmony, which as noted above is a weighted sum of the candidate's constraint violations.

(10) Formula for harmony: $H(x) = \sum_{i=1}^N w_i C_i(x)$

Here, $H(x)$ designates the harmony of candidate x ; w_i is the weight of the i th constraint (weights appear in top row of 9), and $C_i(x)$ is the number of times x violates the i th constraint.

$\sum_{i=1}^N$ denotes summation over all N constraints.

The next step is to negate the harmony values and take e (the base of natural logarithms, about 2.718) to the result. The resulting number, which Wilson (2014) calls eHarmony, is then used in the next stage of the computation.

(11) Formula for eHarmony: $eHarmony(x) = e^{-H(x)}$

The last step is total up the eHarmony of all candidates; this sum is labeled Z . The predicted probability of any candidate is then its share of Z .

(12) Formula for computing probability

a. Computing Z : $Z = \sum_j eHarmony(j)$

b. Computing predicted probability of candidate x : $p(x) = \frac{eHarmony(x)}{Z}$

In our annotated sample tableau in 9, the nasal-substituted candidate thus receives a probability of 0.73, and the unsubstituted one 0.27, not far off from the observed probabilities of 0.67 (60.5/90 words) and 0.33 (29.5/65 words).

The substance embodied in the MaxEnt calculations is as follows. First, negation of Harmony means that it acts as a kind of penalty score, so that heavily penalized candidates get lower eHarmonies and thus ultimately lower probabilities (see fn. 4). Second, constraints with higher weights have more powerful effects, in that candidates that violate them are assigned lower probabilities. Third, as with all forms of Harmonic Grammar (and unlike in strict-ranking Optimality Theory) there is pervasive CUMULATIVITY (Jäger & Rosenbach 2006): violations of two weaker constraints (or multiple violations of a single weak constraint) can outweigh a violation of a stronger constraint. The cumulative property will emerge as crucial in the studies to follow.

Readers often ask for an intuitive explanation of why harmony must be exponentiated (as in 11). Exponentiation performs two functions: it keeps probability from ever going negative (which is logically impossible), and it implements our human intuition that a great deal of evidence is needed when we are aiming at certainty. For example, it can be shown by the equations above that when, in a two-candidate system, candidate C_1 is already at probability .99, then we would have to penalize its opponent C_2 by an additional harmony of 2.3 to get $P(C_1)$ to move up to .999. To reach $P(C_1) = .9999$, we need yet another 2.3 units of harmony. In the medial regions, by contrast, comparable linear shifts of probability come cheaply: it only takes 0.0036 units of harmony added to C_2 to promote $P(C_1)$ from .5 to .5009; and only 0.00036 more units to get $P(C_1)$ up to .50099. Such differences show up geometrically in the sigmoid curves seen below.

THE ROLE OF THE GRAMMAR. What does it mean to formulate a grammar that predicts that 2/3 of the words in a particular class will undergo a phonological process? Individual words have a particular behavior (substitution, nonsubstitution, or variation) that must be encoded somehow in the lexicon, seemingly obviating the need for a grammar. Following earlier work (ZURAW 2000, ZURAW 2010), our model is intended as an account of the PRODUCTIVE BEHAVIOR of Tagalog speakers when they apply η -final prefixes in creating novel forms. In a nonce-probe (wug) test (Berko 1958), Tagalog speakers showed preferences for substitution that reflected the statistics

of their lexicon. The same held for Spanish loanwords that have been incorporated into Tagalog, in effect a real-life wug test over time.

The tendency of productive behavior to match lexical frequencies has been documented in a variety of languages, leading us to give the phenomenon a name, the Law of Frequency Matching (Hayes et al. 2009). In the Hungarian example below, we will model experimental behavior as well as the lexicon.

LEARNING AND THE MAXENT GRAMMAR. The weights used in 9 were fitted using OTSoft (HAYES, TESAR & ZURAW 2014). Weights are fitted so as to maximize the model's predicted probability for the observed data. We set σ^2 , the parameter that determines how willing the weights are to deviate from zero (or another default value), to 10,000, so that there is essentially no penalty on large weights, and the learner's only goal is to fit the data. The weights learned are listed in Table 1, and the resulting predicted probabilities are plotted in Figure 5.

Table 1<INSERT Table 1 HERE>

Figure 5<INSERT Figure 5HERE>

The predicted probabilities should be compared with the observed probabilities given earlier in Figure 4. It can be seen that the MaxEnt grammar grossly captures the observed ceiling and floor effects: consonantal differences are biggest for the intermediate prefixes, and prefix differences are biggest for the intermediate consonants, flattening out for the most extreme consonants on both ends, although the model does not succeed in producing uniformly low nasal substitution rates for /g/. (Fit of this and other models is quantified in Section 2.7.)

MATHEMATICAL BASIS OF THE FLOOR AND CEILING EFFECTS. The reason that MaxEnt produces floor and ceiling effects has to do with the sigmoid shape of candidate probability as a function of constraint-weight difference. For each consonant, there is a certain MARKEDNESS DIFFERENCE between the substituted and unsubstituted candidates—specifically, the weights of the applicable $*[\text{root } \text{NASAL}]$ constraints (which favor the unsubstituted candidate) minus the weights of NASSUB

and *NÇ if applicable (which favor the substituted candidate). In the tableau in 9 for /paŋ-Noun+t ... /, this was $(0.00 + 2.13) - (2.31 + 4.85) = -5.0$ (after rounding). The probability of nasal substitution for each prefix at each consonant is a function of the markedness difference for that consonant, and of the relevant UNIFORMITY constraint's weight. The function is logistic, of the form $p(subst) = \frac{1}{1 + e^{\text{markedness_difference} + \text{weight}(UNIFORMITY)}}$. Details of how this function is derived are shown in 13, for the example consonant /b/.

(13) Probability of nasal substitution for /b/

Exponentiated constraint violations of substituted candidate, divided by sum of exponentiated violations for both candidates:

$$\frac{e^{-\text{weight}(UNIFORMITY) - \text{weight}(*[m/n/ng])}}{e^{-\text{weight}(UNIFORMITY) - \text{weight}(*[m/n/ng])} + e^{-\text{weight}(NasSub)}}$$

Divide top and bottom by the numerator:

$$= \frac{1}{1 + e^{\text{weight}(*[m/n/ng]) - \text{weight}(NasSub) + \text{weight}(UNIFORMITY)}}$$

With the fitted weights in our grammar, the probability function for *maŋ-other*, the most-substituting prefix, is $p(subst) = \frac{1}{1 + e^{\text{markedness_difference} + 0.00}}$ and for *paŋ-res*, the least-

substituting prefix, $p(subst) = \frac{1}{1 + e^{\text{markedness_difference} + 6.01}}$. These two versions of the function

are plotted in Figure 6. Dashed vertical lines show the markedness difference for each consonant. The plot shows that the prefixes are best distinguished—that is, the two sigmoids are the farthest apart—for the intermediate consonants, and begin to collapse together against the ceiling or floor for the consonants at each extreme. As the markedness difference becomes more negative, both

denominators converge towards $1+0$. Thus, the two probabilities converge towards $1/(1+0) = 1$. As the markedness difference becomes more positive, both denominators grow so large that both probabilities are indistinguishably close to zero.

Figure 6<INSERT Figure 6 HERE>

In general, a function of the form $\frac{1}{1+e^n}$ is very close to either 1 or 0 for most values of n . In order to see a value between 0.99 and 0.01, n must be between about -5 and 5, and to see a value between 0.90 and 0.10, n must be between about -2.2 and 2.2. In our example, this means that if either group of constraints is contributing a large negative or positive value to n —such as p 's markedness difference of -7.2—then it will take large opposing value contributed by the other group—UNIFORMITY—to pull the probability of the substituted candidate away from the ceiling or floor. The result is that choice of prefix matters less for the extreme consonants p and g than for the intermediate consonants k and b .

Similar calculations show that consonant differences are greatest for UNIFORMITY constraints with intermediate weights, and smallest for those with extreme weights (see supplemental materials for examples).

In this example, we have two dimensions of variation: constraints relevant to the consonant were grouped into the MARKEDNESS DIFFERENCE, and constraints relevant to the morphology were in the UNIFORMITY family. We have shown that when the difference (in weighted constraint violations) between two candidates is already large in one dimension, the other dimension has little effect. In sum: the ability of MaxEnt to describe across-the-board patterns constrained by floor and ceiling effects is not accidental, but is a general prediction of the model, a direct consequence of the equations it employs. As we will see next, the same is true of Noisy Harmonic Grammar.

NOISY HARMONIC GRAMMAR MODEL. Another method for making Harmonic Grammar probabilistic is Noisy Harmonic Grammar (Boersma 1998a, Boersma & Pater 2013). This system likewise starts by computing a harmony score for each candidate as given above in 10. Instead of

exponentiating and normalizing to obtain a probability, however, NHG simply designates the candidate with the best score as the winner. What makes the model NOISY (i.e. probabilistic) is that at each evaluation time, some Gaussian noise is added to each constraint's weight, potentially changing the winning candidate. A candidate's probability of winning is most easily determined by simulation: run the grammar some large number of times, and observe how often each candidate wins.

An example is shown in 14 for one evaluation instance.⁶ In this example, the winner after noise (substitution) is the same as the winner before noise. The farther apart the harmonies before noise, the less likely it is that noise can overturn the before-noise preference.

(14) Sample Noisy HG tableau—one evaluation

<INSERT Tableau 14 HERE>

Boersma and Pater (2008) have put forth an algorithm similar to Boersma's (1998b) Gradual Learning Algorithm, and have shown that it is effective for learning weights in Noisy HG. We used this algorithm to fit a Noisy HG grammar to the Tagalog data, with the same constraints as in the previous section and using OTSoft (Hayes, Tesar & Zuraw 2014).⁷ The resulting weights (used in the example above) are given in 15, and the model's probabilities are shown in Figure 7.

(15) Noisy HG weights, fitted to Tagalog

<INSERT Table 15 HERE>

Figure 7 <INSERT Figure 7 HERE>

The Noisy HG model captures the ceiling and floor effects, for similar reasons as in the MaxEnt model. To take just the simple case of /p/, the probability of substitution is the probability that the weights-plus-noise of $*_{\lceil \text{root } m/n/\eta}$ and UNIFORMITY are together smaller than those of *NÇ and NASSUB, as in 16.

- (16) probability of nasal substitution for /p/ = probability that the following inequality holds
- $$\text{weight}(*[\text{root } m/n/\eta]) + \text{noise} + \text{weight}(\text{UNIF}) + \text{noise} < \text{weight}(*\text{N}\underset{\circ}{\text{C}}) + \text{noise} + \text{weight}(\text{NASSUB}) + \text{noise}$$
- i.e.

$$\text{noise} + \text{noise} - \text{noise} - \text{noise} < \text{weight}(*\text{N}\underset{\circ}{\text{C}}) + \text{weight}(\text{NASSUB}) - \text{weight}(*[\text{root } m/n/\eta]) - \text{weight}(\text{UNIF})$$

In other words, the probability of nasal substitution on /p/ is the probability that the sum⁸ of four independent, Gaussian random variables is less than a certain number (which will vary according to which UNIFORMITY constraint is in play). The sum of four independent Gaussians is itself a Gaussian, and the probability that it is less than some number—known as the cumulative distribution function—is sigmoid (Peters 2002: 4-5). Figure 8 shows this sigmoid function for /p/, and the analogous one for /k/. We can see that /p/'s probability of substituting clings to the ceiling until high values of UNIF (about 10), and that at hypothetical even higher values of UNIF, both consonants' rates hit the floor. Vertical lines indicate the actual UNIFORMITY weights for our six prefixes.

Figure 8<INSERT Figure 8 HERE>

As in the MaxEnt model, it is only when the two candidates (substituted and unsubstituted) are otherwise close competitors that prefix or consonant differences can be seen. When, for example, faithfulness is weighted too low to resist $*\text{N}\underset{\circ}{\text{C}}$ and $*\text{NASSUB}$, giving a large advantage to the substituted candidate, little difference between /p/ and /k/ is observed. Only when the substituted and unsubstituted candidates' scores are otherwise reasonably close do the weights of $*[\text{root } n/\eta]$ and $*[\text{root } \eta]$ affect the outcome enough for a difference between /p/ and /k/ to emerge.

To sum up: Noisy Harmonic Grammar, just like MaxEnt, naturally generates sigmoid curves as a consequence of its basic mathematical form. These sigmoids, though not identical to

the MaxEnt sigmoids, serve the same function in accounting for across-the-board effects, overridden by floor and ceiling effects.

2.6 SOME UNSUCCESSFUL MODELS. We now consider some models that are less successful in capturing the Tagalog data.

STOCHASTIC OT MODEL In contrast to our two Harmonic Grammar models, which employ constraint weighting, a Stochastic OT grammar (Boersma 1998a) is a probability distribution over strict-ranking (classical) OT grammars. These probability distributions are quite restricted: they must be expressible by assigning each constraint a RANKING VALUE, such that the probability of a given constraint ranking is the probability that the ranking values, after adding noise, fall into the corresponding order. In other words, at evaluation time, Gaussian noise is added to each constraint's ranking value, the constraints are ranked according to the resulting numbers, and strict-ranking OT applies as usual. Two constraints whose ranking values are far apart effectively have a fixed ranking, and those that are closer together vary appreciably in ranking.

The Gradual Learning Algorithm (Boersma 1998a) can be used to learn Stochastic OT grammars, although as Pater (2008) discusses, the algorithm is not guaranteed to converge and even if it does converge the result is not guaranteed to be the Stochastic OT grammar that best matches the data. To help address this problem, we used a modified version of OTSoft (Hayes, Tesar & Zuraw 2014) to run 999 learning simulations.⁹ We also ran 999 simulations using a slightly different version of the GLA incorporating an update rule devised by Magri (2012) in response to Pater's (2008) criticism of the original update rule. We picked the grammar that was the best fit (assigns the greatest log likelihood to the observed data) out of these, though it is still possible that a better-fitting grammar exists.

The ranking values for the best grammar, which was one of the Magri-fitted grammars, are listed in Table 2 and its predicted probabilities, derived by simulation in OTSoft, are plotted in Figure 9.

Table 2<INSERT Table 2 HERE>

Figure 9<INSERT Figure 9 HERE>

The strongly negative weights (and one weight much lower than the rest) indicate that the learner is failing to converge; in particular, the weight of $*_{\text{root}} \text{m/n/}\eta$ continues to be pushed downwards, even after its weight is so low that it will never outrank a competing constraint. The extremely low ranking of $*_{\text{root}} \text{m/n/}\eta$ —too low for it ever to be decisive—means that there is virtually no difference between labials and coronals: p and t/s have nearly the same predicted behavior, and, more problematically, so do b and d , although they are quite different in the training data. There is little difference among the voiceless consonants at all, because $*_{\text{root}} \eta$ is not ranked high enough to compete with $*\text{NC}_{\text{C}}$ more than slightly; the competition is almost entirely between $*\text{NC}_{\text{C}}$ and UNIFORMITY constraints, none of which care about place of articulation. If $*_{\text{root}} \text{m/n/}\eta$ were ranked higher, however, then the model would do even worse at capturing p 's ceiling-wards tendency, and if $*_{\text{root}} \eta$ were ranked higher, then it would pull down substitution rates for k , not so much for the less-substituting prefixes, but for the most-substituting prefixes, whose ranking values are closest to $*_{\text{root}} \eta$'s; this would be a poor match to the training data. More generally, the model encounters repeated contradictions in finding a suitable placement for constraints on the ranking scale; they cannot be in two places at once, as we now discuss.

WHY DOES THE STOCHASTIC OT MODEL FAIL FOR TAGALOG? In certain instances it is possible to construct a Stochastic OT grammar that captures a family-of-sigmoids shape, but only with a set of constraints that is otherwise undesirable. We illustrate this by adopting an alternative constraint set for Tagalog that gives up on the goal of characterizing the voicing and place effects as interacting dimensions, and rather treats each voice/place combination as independent. We will first show that this constraint set succeeds, and then show why the preferred set of constraints fails: For our alternative constraint set, we simply adopt one markedness constraint demanding nasal substitution for each underlying consonant, $*\text{N+P}$, $*\text{N+T/S}$, and so on. Training the GLA with these constraints on the observed data (using the same procedure of selecting the

best-fit model from 999 runs each, with and without Magri's update rule), we obtain the ranking values shown in Table 3 and the predictions seen in Figure 10.

Table 3<INSERT Table 3 HERE>

Figure 10<INSERT Figure 10 HERE>

The grammar works because each data point is modeled as the interaction of just two opposing constraints. Convergence toward zero or one at the extremes is possible if the relevant markedness constraint is either extremely weak (*N+g) or quite strong (*N+p). In the middle, *N+b has an intermediate ranking value, producing a wide range of probabilities that it can outrank each UNIFORMITY constraint.

But our straw-man grammar misses the generalizations about voicing and place, which have typological support (Newman 1984, Blust 2004, Pater 2001, Zuraw 2010). It works only because each dimension—prefixes and consonants—lacks any internal structure. If we introduce internal structure, through constraints like *NÇ (which is relevant for multiple consonant places) or *[_{root}ŋ (relevant for both /k/ and /g/), then the Stochastic OT architecture can no longer fit a family of converging sigmoids, as we saw in Figure 9.

The failure can be diagnosed with standard forms of ranking argumentation. We will build up a grammar to fit observed data points one by one, and arrive at a ranking paradox. First, to ensure near-zero nasal-substitution rates for /g/ with all prefixes, we need *[_{root}ŋ to have a ranking value far above that of NASSUB (the only constraint favoring substitution for /g/). A 7-point difference will keep substitution rates for /g/ below 1% (matching the empirically observed value).

This difference is shown in the grammar on the left in Figure 11, where NASSUB's ranking value is 93 and *[_{root}ŋ's is 100. The diagram is similar to a Hasse diagram used to depict the known ranking relation in a non-probabilistic OT grammar. The difference is that in a typical Hasse diagram, a line between two constraints indicates that the constraint on top outranks the one on the bottom, and the length of the line is irrelevant. Here, such a line indicates instead a ranking tendency, and the longer the line, the stronger the tendency.

Consider next /b/. It has a high rate of substitution with *maŋ-other*, so NASSUB, the only constraint favoring substitution of /b/, must tend to outrank both $*_{[\text{root } m/n/\eta]}$ and UNIF-*maŋ-OTHER*. A difference of 4 points (as shown in Figure 11) means that each substitution-penalizing constraint has an 8% chance of outranking NASSUB, and the combined probability that one or the other outranks it is a little higher, 13%, giving /*maŋ-other*+b ... / an 83% rate of substitution.

As for /b/ with *paŋ-noun*, we want a lowish rate of substitution, around 33%. $*_{[\text{root } m/n/\eta]}$ can't do the job of enforcing this low rate, because we've just established that it is ranked well below NASSUB. The task therefore falls to UNIF-*paŋ-NOUN*; we can give it a ranking value of 94, so that NASSUB outranks it 36% of the time (the actual rate of substitution will be a little lower, because even when UNIF-*paŋ-NOUN* doesn't outrank NASSUB, occasionally $*_{[\text{root } m/n/\eta]}$ will do so).

Next we consider the input /*maŋ-other*+k ... /. We need a high substitution rate despite $*_{[\text{root } \eta]}$'s objection. NASSUB can't be responsible, because at six points below $*_{[\text{root } \eta]}$, it is unlikely to outrank $*_{[\text{root } \eta]}$. The only constraint besides NASSUB that favors substitution is $*\text{NC}\text{̥}$, so it must have a ranking value substantially higher than $*_{[\text{root } \eta]}$'s (and also higher than UNIF-*maŋ*'s, but this is ensured transitively by what we've already established). This is again shown in the grammar on the left in Figure 11.

Figure 11<INSERT Figure 11 HERE>

But now we have predicted, incorrectly, that all nasal-substitution rates for /p/ and /k/ will be very close to 100%, because $*\text{NC}\text{̥}$ nearly always comes out as the top-ranked relevant constraint. The contradictory ranking fragment needed to fit voiceless consonants with the *paŋ-noun* prefix is shown on the right in Figure 11. We need UNIF-*paŋ-NOUN* to have a good chance of outranking both $*\text{NC}\text{̥}$ and NASSUB, which happens if we increase its ranking value to 103. (This yields a nasal-substitution rate of around 85% for /*paŋ-noun*+p ... /. We also have to promote $*_{[\text{root } \eta]}$ high enough to produce a lower rate for /*paŋ-noun*+k ... /, about 60%).

Thus we have the probabilistic version of a ranking paradox: UNIF-*paŋ-NOUN* can't be in two places at once. On the one hand, in the grammar on the left, its ranking value must be just a little above NASSUB's (to model /*paŋ-noun*+b ... /), and NASSUB must be far below $*\text{NC}\text{̥}$ (to model /*maŋ-other*+k ... /)—but on the other hand, in the grammar on the right, its ranking value should be close to $*\text{NC}\text{̥}$'s, to model voiceless stops with *paŋ-noun*.

To summarize, if we model complexity of the consonant dimension, using constraints that were previously proposed (Zuraw 2010) on cross-linguistic, language-internal, and phonetic grounds, Stochastic OT is not able to capture the quantitative pattern of Tagalog nasal substitution. In other words, the constraint set matters. With just the right constraints—in this case one for each consonant—it is possible to produce a family of sigmoids in Stochastic OT. MaxEnt and Noisy Harmonic Grammar, by contrast, produce a family of sigmoids regardless of constraint set. They are persistently sigmoidal, where StOT is only sporadically sigmoidal. If our three case studies are typical, this makes MaxEnt and Noisy HG more promising as quantitative models of constraint interaction.

ANTTILIAN PARTIAL ORDERING. As in Stochastic OT, in Anttila's theory of partial ordering (Anttila 1997a; Anttila 1997b), a grammar is a probability distribution over total rankings. The distribution is defined by a partial ordering of constraints. All total orderings that are consistent with that partial ordering are taken as equally probable. For example, if the grammar is the partial ordering $A \gg \{B, C\} \gg D$, then the total rankings $A \gg B \gg C \gg D$ and $A \gg C \gg B \gg D$ are each used 50% of the time. If those two rankings produce different output candidates, then each of those candidates is predicted to occur 50% of the time.

In the original research that established the theory, Anttila generally adhered to a more restrictive version of partial ordering in which the ordering is STRATIFIED (Boersma 2001): constraints within the same stratum are ranked freely, but rankings across strata are fixed.¹⁰ In the example just given there are three strata: top-ranked $\{A\}$, middle-ranked $\{B, C\}$, and bottom-ranked $\{D\}$. We adopt stratification here because (unlike for partial ordering) there is an accessible learning procedure for stratified grammars.¹¹

Specifically, stratified grammars can be learned using the GLA by setting the learning increment to a high number such as 20. This means that all constraints will either have the same ranking value (be in the same stratum) or be at least 20 units apart in ranking value, which is tantamount to an absolute ranking. As before, we ran the GLA 999 times and selected the grammar with the best fit to the data. That grammar is given in Table 4; it has seven strata. We also include ranking values from which the strata were inferred (these values have no status in

this theory, but were included to illustrate how this is a special case of a Stochastic OT grammar.) The grammar's predictions are plotted in Figure 12.

Table 4<INSERT Table 4 HERE>

Figure 12<INSERT Figure 12 HERE>

For voiceless consonants *p*, *t/s*, *k*, the choice is made in the first stratum: if the prefix is *paŋ-noun* or *paŋ-res*, then there is a 50-50 competition between UNIFORMITY and *NÇ, with a resulting 50% rate of nasal substitution. For other morphology, *NÇ prevails, and the substitution rate is 100%. As a result, there are no differences at all among voiceless consonants, and only two different rates of nasal substitution observed.

The modeling of the voiced consonants is more successful, though only three classes of prefix are distinguished. Voiced consonants *b*, *d*, *g* have a zero rate of substitution for *paŋ-noun* and *paŋ-res*, chosen in Stratum 1. For *g* with other prefixes, the choice is made in the next stratum by *[root ŋ], yielding 0% substitution. For *b*, there is 50-50 competition in Stratum 3 for *maŋ-RED-*, and 100% substitution enforced by NASSUB at Stratum 3 for the remaining morphology. The most complex case is *d*, where there are two relevant constraints at Stratum 3 for most prefixes, producing 50-50 competition between NASSUB and *[root n/ŋ]. But for *maŋ-RED-*, there are two constraints favoring non-substitution against one constraint favoring substitution; whichever constraint is ranked at the top will make the decision, and each has an equal probability of being ranked at the top, so the resulting substitution rate is 33%.

In sum, the best stratified Anttilian model that could be learned was highly defective: it fails to distinguish the voiceless consonants from one another and it distinguishes at most three prefixes. We have not located ANY circumstances under which the Anttilian model is able to generate a family of sigmoids, although we cannot rule out the possibility that some such circumstances may exist.

More generally, with our constraint set, it is not possible for substitution rates to converge towards 1 for /p/--that is, within the voiceless consonants, it is not possible for prefix differences to be smallest for /p/ and biggest for /k/. The prefix differences can be identical across the

voiceless consonants, as in the best model we found, or they can be largest for /p/ and smallest for /k/, if some or all of the *[_{root} NASAL constraints are promoted into the top stratum.

DECISION-TREE MODEL. Our next model is based on a radically different premise from the others: suppose that the Tagalog pattern is the consequence of sequenced MODULES, as in the classical feed-forward model of generative grammar. The modules themselves may well involve interacting constraints, but the two intersecting constraint families would reside in separate components.

We can flesh out this idea as follows. A morphology/syntax module first probabilistically assigns structure, which is then fed into a phonology module. Schematically, the structures assigned would be [prefix [root]_{domain}] contrasting with [prefix root]_{domain}, with the details depending on the particular theoretical framework employed (see Gref 2008 for a concrete proposal for a cognate prefix). [prefix [root]_{domain}] is a LOOSE structure (only the root is within a certain domain) and [prefix root]_{domain} is a CLOSE structure (prefix and root are together in that domain).

Let's further suppose that if a word is assigned the loose structure, there is no possibility of nasal substitution in the phonological module—the prefix and stem fail to be together in nasal substitution's domain of application—but if a word is assigned the close structure, then the phonology probabilistically decides whether to apply nasal substitution.

We can describe this sequence of probabilistic choices using a DECISION TREE (Breiman et al. 1984), illustrated in Figure 13 for a hypothetical prefix and consonant.

Figure 13<INSERT Figure 13 HERE>

Assuming this tree, a novel word's probability of undergoing nasal substitution is its probability of getting the close structure (which depends on the prefix), multiplied by its probability of then undergoing nasal substitution in the phonology (which depends on the consonant).

We implemented a decision-tree model of the Tagalog data. To give the model the benefit of the doubt, we allowed it to assign independent probabilities of substitution for each consonant, ignoring the general patterns based on voicing and place; the model also included a separate probability of close structure for each prefix. To obtain the best-fit probabilities, we

used the `optim()` function in R (R Core Team 2014), with the objective of maximizing likelihood (the probability that the model assigns to the observed data). The best-fit probabilities were as follows.

Table 5<INSERT Table 5 HERE>

As Figure 14 shows, this model fit the data poorly. (Only four lines can be seen, because the top three prefixes were assigned identical probabilities of close structure.)

Figure 14<INSERT Figure 14 HERE>

The model succeeds in having probabilities converge near zero at the right end of the consonant scale. This is because the fitted probability of substitution, given close structure, for /g/ is low, 0.01, imposing an upper limit on how often /g/ can undergo substitution: multiplying such a low number by even *may-other*'s high probability of close structure (1.00) still results in a very small probability (0.01). But the model necessarily fails to achieve convergence near 1 at the left end of the consonant scale. The prefix *pay-res.* has the lowest probability of close structure, 0.27, so that no consonant can ever achieve a rate of substitution higher than 0.27 with that prefix. Even though /p/'s probability of substitution, given close structure, is high (1.00), multiplying by 0.27 still results in a low probability (0.27). Multiplying two probabilities can never produce a probability larger than either of the two. Thus differences among prefixes become AMPLIFIED at the left end of the scale, rather than reduced towards 1 as in the observed data.

Because the decision-tree model cannot generate the desired family-of-sigmoids pattern, we conclude that this modular account is not plausible for Tagalog. Instead, the constraints that depend on the morphology and those that depend on the consonant must interact within a single module. We see this result as support for an emerging view that phonology and morphosyntax are not in a strictly modular, strictly feed-forward relationship (see Shih 2014 for novel data and arguments and a literature review).

2.7 COMPARISON OF MODELS. To summarize the results in this section, we give a quantitative comparison of the models' fit to the Tagalog data in Table 6.¹² To provide a baseline for comparison, we also include a PERFECT MODEL, which simply matches the observed probability of nasal substitution for each consonant/prefix combination. No model that discriminates only consonants, prefixes, and/or their combinations could perform better than this. Our measure of fit is log likelihood: we sum the natural-log probability that the model assigns to each data point. The closer to zero this value is, the more accurate the model.

Table 6<INSERT Table 6 HERE>

We include a column in Table 6 giving the number of free parameters in each model. All else being equal, the more free parameters a model has, the more able it is to fit a variety of data, and the less we should be impressed by its log likelihood. The four OT-based models all have 11 free parameters (one weight or ranking value for each constraint); the decision tree has 12 (one probability for each consonant and one for each prefix); the 'perfect' model given for comparison simply lists the correct rate for each of the 36 cases (6 consonants times 6 prefixes). This is not a perfect measure of model complexity, because all else is NOT equal. For example, Stratified Partial Ordering has the same number of free parameters as Stochastic OT, but the values of those parameters are not as free, because they are constrained to be multiples of 20. Constraint structure can also matter: the straw-man model that we created in Table 3 with a separate constraint for each consonant (log likelihood: -269.55) has just 12 constraints—only one more than the other constraint models—but the model is far less constrained than its number of parameters suggests, because each consonant gets its own constraint rather than being subject to the same constraints as consonants that share its voicing or place.

2.8 TAGALOG CONCLUSIONS. We have shown in this section that Tagalog nasal substitution requires a model where each dimension—consonant or prefix—has its main effect in the intermediate range of the other. At the extremes of each dimension, behavior hits a floor or ceiling. Graphically, this appears as a family of sigmoid curves. The two varieties of Harmonic

Grammar we examined, MaxEnt and Noisy Harmonic Grammar, succeed in capturing this pattern.

The other models we examined were far less successful. With the constraint set used, models based on strict ranking (Stochastic OT, partial ordering) failed to generate sigmoid curve families. With a different constraint set — abandoning crucial generalizations about voicing and place — Stochastic OT can succeed. Multiplying two probabilities, as in the decision tree model, cannot generate a sigmoid curve family.

The Tagalog case involves a morphological and a phonological dimension. In the remainder of the paper we examine two more cases: French liaison/élision, where morphosyntax interacts with lexical marking, and Hungarian vowel harmony, where purely phonological dimensions interact.

3. FRENCH LIAISON/ELISION: LEXICAL MARKING VS. MORPHOSYNTAX. Our second case study examines two-word sequences in French, abbreviated Word1+Word2, where Word1, one of various function words or adjectives, has two allomorphs, chosen according to phonological and lexical properties of Word2. We examine the interaction pattern between constraints that distinguish the Word1s and constraints that distinguish the Word2s. Although this is well-plowed ground, we do make novel empirical claims about differences among Word1s.

3.1 LIAISON AND ELISION. The alternations in Word1 are cases of French’s well-known liaison and elision phenomena. Some examples are shown in Table 7. The default allomorph of the Word1s under consideration ends in a vowel, and is used in citation form or when followed by a Word2 that begins with a consonant. When Word2 begins with a vowel, Word1 has a different allomorph that is consonant-final, either through addition of a final consonant (liaison¹³) or deletion of the final vowel (elision).

Table 7<INSERT Table 7 HERE>

This simplified, non-probabilistic description is easily captured in any of the grammar models discussed above for Tagalog, given suitable constraints. The challenge comes in capturing variation, which we will demonstrate below exists.

Before moving on, we give for reference the full list of Word1s whose two allomorphs are spelled differently, in Table 8. Because we use written data (discussed below), we cannot study any Word1s whose allomorphs are spelled the same, such as *mon* ‘my’-*masc.*, which alternates between [mō] and [mōn]. The Word1s are a variety of function words, and a small number of adjectives that, counter to the general pattern in French, can precede the noun.

Table 8<INSERT Table 8 HERE>

3.2 LEXICALLY SPECIFIC BEHAVIOR. Idiosyncratically, some vowel-initial Word2s take the preconsonantal allomorph of Word1.

(17) Vowel-initial Word2s that behave as though consonant-initial

<i>la hache</i>	[la aʃ]	‘the axe’
<i>du haricot</i>	[dy aʁiko]	‘of the bean’
<i>le uhlan</i>	[lə ylã]	‘the uhlan (central/east European lancer)’

Most of these words are spelled with an initial silent¹⁴ <h>, known as *h aspiré* (‘aspirated *h*’), and the word’s behavior is often a vestige of an earlier or foreign-language pronunciation that did have an initial consonant. Only a portion of words spelled with <h> behave this way; the rest behave as ordinary vowel-initial words. In keeping with the analysis to be adopted below, which relies on syllable alignment, we coin the term ALIGNANT to refer to words like those in 17, and NON-ALIGNANT to ordinary vowel-initial words. In brief, alignant words prefer to begin with a syllable boundary—that is, to be aligned at their beginning with a syllable boundary; one way to enforce this is to use the preconsonantal allomorph of Word1.

Words that begin with a glide likewise behave idiosyncratically as either alignant or non-alignant in their choice of Word1.

Table 9<INSERT Table 9 HERE>

3.3 PREVIOUS CLAIMS OF VARIATION. If a word begins with a vowel or glide, French speakers must memorize whether it behaves as alignant. Spelling is but an approximate guide: vowel-initial words spelled with <h> can easily fall into either category, and a few alignant words are not spelled with <h>. Glide-initial alignant words tend to be spelled with <y> or <w> (rather than *(h)i*, *(h)ou*, etc.), but again the correlation is imperfect.

It is not surprising, then, that there is widespread variation, as prescriptive grammarians and linguists have long noted. To avoid getting bogged down in examples and citations, we have placed them in the Appendix.

Many authors have claimed that there is not only across-speaker disagreement as to lexical entries, but also intra-speaker variation. Most important for our purposes, however, are the numerous claims in the literature of tendencies for a given Word2 to behave differently depending on Word1 (again, see the appendix). For example, it has been claimed that *oui* ‘yes’ usually behaves as alignant, but tends to behave as non-alignant with *que* ‘that’. Generalizations about such differences are rare in the literature (e.g. claims about *de* in general vs. *que* in general), though we will claim below that such generalizations can be made.

The distinct behavior of different Word1s becomes particularly evident when a Word2 changes its alignant status over time. The word *hiatus* ‘hiatus’ has shifted in recent decades from non-alignant to alignant. The plot in Figure 15 shows rates of use over time—as a percentage of all text—for *le/l’+hiatus* and *de/d’+hiatus* in the Google n-grams corpus using its visualization interface (see section 3.5).¹⁵ Alignant *de hiatus* overtakes *d’hiatus* around 1950, but *le hiatus* doesn’t overtake *l’hiatus* until about 1970. This is in line with our finding below that the Word1 *le* is in general more likely than *de* to treat the following word as alignant.

Figure 15<INSERT Figure 15 HERE>

3.4 ANALYTIC APPROACH. We will not attempt a full review of the extensive and long-running literature on lexical effects in French phrasal sandhi. The main approaches are, as surveyed by Tranel (1995), representational, diacritic, and prosodic; we will adopt a prosodic approach.

In representational approaches, the lexical entry of an alignant word begins with some kind of consonant, such as an [h] or [ʔ], that is deleted late (e.g. Schane 1968, Selkirk & Vergnaud 1973, Dell 1973, Schane 1978, and one option considered in Boersma 2007), or an empty consonantal slot (e.g. Clements & Keyser 1983). For example, ‘tree’, which behaves as a normal vowel-initial word, would have the lexical entry /aʁbɛ/, and so select *l’*, *d’*, and so on; but ‘bean’ would be /haxiko/, selecting *le*, *de*, and so on. Glide-initial words can be treated in the same way (alignants begin with /h/ or some other consonant, and non-alignants begin with an underlying high vowel, e.g. Martinet 1933), or by treating alignant words as beginning with an underlying glide and non-alignant as beginning with an underlying high vowel (e.g. Kaye & Lowenstamm 1984, Clements & Keyser 1983).

The motivation for employing an abstract consonant is that alignant words largely behave as though they were consonant-initial. However, several authors have pointed out that this is not entirely the case. The best-known challenge involves schwa/zero alternations (see appendix for full examples). Unlike either an ordinary consonant-initial word or an ordinary vowel-initial word, an alignant word optionally causes a schwa to appear before it (the schwa either is inserted or resists deletion, depending on the underlying form assumed): for example, alignant [ynə] *hache* ‘an axe’, vs. consonantal [yn] *tache* ‘a spot’ and non-alignant [yn] *honnête femme* ‘an honest woman’ (Grammont 1914: 124, Schane 1968, Selkirk 1972: 329, 379; Dell 1973: 84-93; Fouché 1956: 135-139; Grevisse & Goosse 2011: 55, 63). Tranel (1981) argues that there is an optional rule of schwa insertion. The representational analysis provides no explanation of why schwa insertion should occur, but as we will see, it is entirely expected under the alignment account.

Diacritic approaches mark alignant words as exceptionally undergoing, resisting, triggering, or not triggering rules. (e.g. [-LIAISON]). Tranel 1981 advocates diacritics because they provide the needed flexibility to mark some words as, for example, triggering liaison but not schwa insertion.. But this misses the overall bundling tendency of properties. Most crucially, the diacritic approach fails to capture the prosodic commonality of the rules in question, unlike the prosodic approach, to which we now turn.

The prosodic approach has a long history (see Frei 1929, Damourette & Pichon 1930, Malmberg 1943 in Appendix). The key idea is that alignant words carry a requirement to begin with a syllable boundary. In the analysis developed below, alignant words are characterized by being subject to the alignment constraint ALIGN(Word, Left; Syllable, Left): the beginning of the word must also be the beginning of a syllable (Prince & Smolensky 1993/2004, McCarthy & Prince 1993). Tranel (1981: 299) discusses this possibility in pre-OT terms; OT analyses along these lines have been pursued by Tranel & Del Gobbo (2002) and Boersma (2007).

The prosodic approach has three benefits. First, as compared to the diacritic approach, it explains why liaison, elision, resyllabification, and schwa insertion strongly tend to pattern together. In essence, the prosodic approach substitutes a single diacritic—propensity to align—for a set of independent diacritics.

Second, as compared to the representational (latent-consonant) approach, the prosodic approach explains why alignant words act as consonant-initial in most but not all respects. Acting as though consonant-initial by suppressing liaison maintains a syllable boundary before the word: *beau hérisson* [bo.e.ʁi.sõ] vs. **bel hérisson* [bɛ.le.ʁi.sõ] ‘nice hedgehog’; so does suppressing elision: *le hérisson* [lə.e.ʁi.sõ] vs. **l’hérisson* [le.ʁi.sõ]. As Cornulier (1981: 184-185) points out, glide formation is also suppressed before alignant words just as before consonants, as in *qui hasarde* [ki.a.zaʁd], rather than *[kja.zaʁd] ‘who hazards’, again maintaining the syllable boundary. Omitting schwa, however, as before a consonant-initial word, does not maintain a syllable boundary, as illustrated in 18: *une hache* ‘an axe’, if pronounced without a schwa, must violate either ALIGN or a prohibition on having a syllable boundary between a consonant and vowel, *C.V (see Tranel & Del Gobbo 2002). Including a schwa avoids both of these problems. There is thus a competition, leading to variation, among ALIGN, *C.V and DEP-ə/*ə (see Gabriel & Meisenburg 2009 for some data on this variation).

(18) Optional schwa before alignant words

<INSERT Tableau 18 HERE>

Another way that alignant words do not behave like normal consonant-initial words is a paradigm gap claimed by Cornulier (1981): a consonant-final prefix cannot combine with an

alignant stem: **ex-héros* ‘ex-hero’, **mal-hanté* ‘mis-haunted’, **super-hâlé* ‘super-tanned’, **sur-housse* ‘over-cover’. Although tokens of some of these can be found on the web, Cornulier’s intuition of a gap presumably indicates that they are degraded. According to Cornulier, the reason for these gaps is the misalignment of morpheme boundary and syllable boundary. These prefixes freely combine, by contrast, with non-alignant and consonant-initial stems.

The third benefit of the prosodic approach, as compared to either the diacritic or the representational approaches, is that it ties in well with a set of observations scattered throughout the literature on ‘setting apart’ certain types of words in French (Ehrhard 1965, Gaatone 1978, and Cornulier 1981 draw together many of these observations; see the appendix for full examples and citations). First, any vowel-initial word can be treated as alignant when used metalinguistically, as in *diminutif de aube* ‘diminutive of dawn’ (Grevisse & Goosse 2011: 64). Second, archaic, technical, and foreign words have a greater tendency to be alignant. And third, short words—whose recognizability in running speech is under greater threat—are more likely to behave as alignant. The length effect also interacts with the other two factors. For example, metalinguistic uses and foreign words are especially likely to be alignant when monosyllabic.

To sum up these effects, there are various pragmatic or processing reasons for a word to need to be set apart—that is, to begin with a syllable boundary. What is special about alignant words is that they share this behavior as a pure lexical idiosyncrasy.¹⁶ Malmberg (1943: 44) writes, ‘The *h* aspiré is a sort of halt that one makes before certain words because one has once learned—at school or elsewhere—that they are dangerous and that one must be careful’.

Thus, we follow earlier researchers in treating alignant words as subject to an alignment constraint. We go beyond this work in proposing that words occupy a spectrum from alignant to non-alignant, so that the strength of ALIGN differs word by word. To implement this, we break ALIGN into a small number of copies, with different strengths and associated to different sets of words, rather than giving each word its own constraint. We propose that phonological theories should incorporate a notion of LEXICAL PROPENSITY, defined as the affiliation of individual lexical items with particular copies of a constraint. The six copies of UNIFORMITY, used for different prefixes in Tagalog above, could likewise be seen as instances of (sub)lexical propensity. In a Word1+Word2 combination like *le homard* ‘the lobster’, Word1 will bear its own constraint (favoring citation form [lə]), and Word2 will be subject to one of the ALIGN constraints, reflecting its own propensity to align.¹⁷

We will not attempt to account for specific combinations with lexicalized behavior (see appendix)—they will have the effect of noise in our data—but we assume them to be listed as separate lexical entries. We assume that the child acquiring French begins by listing multi-word chunks such as *fromage d'Hollande* ‘type of cheese’ (cf. *de Hollande* ‘of Holland’, which shows that *Hollande* is typically non-alignant), and later begins to extract generalizations about each Word1 and Word2, but with lexicalized chunks still able to override productive combination.

3.5 DATA. To investigate more systematically how alignancy behavior differs according to Word1 and Word2, we looked at a large set of potentially-varying Word2s. We used the Google n-grams corpus for French (Michel et al. 2011, Lin et al. 2012), which includes only sequences that occur at least 40 times in the Google Books corpus; we limited results to publications since 1900. We identified 358 Word2s of interest in the corpus: words with identifiable part of speech and gender (through dictionary look-up); that occur in the corpus with two or more of our target Word1s (that are appropriate for Word2’s part of speech and gender); and are either spelled with initial <h>, pronounced as glide-initial, or listed in dictionary and other prescriptive sources as alignant or varying.

To get a general feel for the data, we first plot a histogram of the mean rate at which each Word2 behaves as non-alignant (averaging over all the Word1s that it happens to occur with). Although rates cluster near 0 and 1, a healthy number of Word2s falls somewhere in between.

Figure 16<INSERT Figure 16 HERE>

The challenge for plotting Word1+Word2 behavior by Word2’s degree of alignancy is that the degree to which a Word2 is alignant can only be observed empirically; there is no a priori way to determine it. To assign each Word2 an empirical degree of alignancy, we fitted a mixed-effects logistic regression model (using function `glmer()` from package `lme4` (Bates et al. 2014) in R (R Core Team 2014). Each observation is a Word1+Word2 sequence; the dependent variable is the rate at which Word1 in that sequence undergoes liaison/elision (i.e. behaves as though Word2 is non-alignant). Only the best-attested and least-problematic Word1s were used: *au/à l’*, *la/l’*, *de/d’*, *le/l’*, *du/de l’*, *beau/bel*, *vieux/vieil*, *ma/mon*. The independent variables are

the identity of Word1 and the log frequency of the Word1+Word2 combination, and each Word2 receives a random intercept. It is unusual to use logistic regression with a dependent variable that is a rate ranging from 0 to 1 (although these rates do cluster strongly around 0 and 1), but the more usual approach, treating each token as an observation with an outcome of aligned or non-aligned, produced models that failed to converge. For comparison, we also performed beta regression (in SAS with the GLIMMIX procedure [SAS Institute Inc 2014], to allow random effects, as this was not available in R), and results were quite similar.

Following Fruehwald (2012), we treat the random intercept that the model assigns to each Word2 as an index of its alignment propensity; positive is non-alignant and negative is alignant. Examples for one slice of the alphabet are given in Table 10, along with, for comparison, each word's normative category and its rate of non-alignant behavior averaged across all Word1s that it happens to occur with.

Table 10<INSERT Table 10 HERE>

This single measure of a Word2's propensity to behave as non-alignant collapses possible phonological sources of Word2 differences, such as vowel quality; non-phonological systematic sources, such as propensity for metalinguistic use; and true item-specific idiosyncrasy.

For the Word1 dimension, we examine the results of the regression model to compare Word1s' tendencies to appear in their preconsonantal/isolation allomorph. Post-hoc pairwise comparisons on the two regression models find three groups of significantly different Word1s: *beau/bel*, *ma/mon*, and *vieux/vieil*, with the highest non-alignment rates; *du/de* and *le/l'* with the lowest, and *au/à l'*, *de/d'*, and *la/l'* in the middle.¹⁸ We are agnostic on why Word1 differences should exist, but we speculate they arise from a variety of sources: syntactic closeness of Word1 and Word2, degree to which Word1+X tends to be a lexicalized unit, and each Word1's arbitrary, idiosyncratic tendency to appear in its citation form versus in its elision or liaison form. Our constraint model below will subsume any syntactic or processing systematicity under each Word1's individual tendency towards non-alignant behavior.

We now have, just as in Tagalog, two cross-cutting dimensions: Word2 alignancy and Word1 identity. We combine the Word1s into three groups according to their coefficients in the regression model (high, medium, and low), as discussed above. We divide the Word2s into five

groups of equal size, along the alignancy continuum, using their random intercepts in the regression model. This grouping facilitates interpretation by providing a reasonable number of items in each Word1-group/Word2-group combination. (Two bins are fairly small, with 9 and 18 items, but the rest are more robust, ranging from 33 to 172 items.) We plot alignment propensity across these two dimensions, as in Figure 17.

Figure 17<INSERT Figure 44 HERE>

Although the differences among Word1s in French are smaller than the differences among prefixes in Tagalog, we see the same basic pattern. At extreme values of Word2 alignancy (1 and 5), Word1 differences disappear: all items hit a floor or ceiling; but in between, Word1 differences are substantial, from about 20% non-alignant behavior to 80% for the middle group of Word2s. Because the Word1 differences are small, and we see only minor floor and ceiling effects in this dimension: the difference between the two lowest Word2 alignancy categories, 1 and 2, is smaller for the lower two Word1 categories, B and C, than for the highest category, A, suggesting a floor effect for B and C; and the difference between Word2 categories 4 and 5 is greater for the lowest Word1 category, C than for A and B, suggesting a ceiling effect for A and B.

3.6 CONSTRAINTS FOR OT-BASED MODELING. Our constraint models all use the same constraint set, listed in Table 11. There is just one constraint encouraging élision/liaison (NOHIATUS). The rest encourage the citation form of Word1, either directly (the Word1 family USEAU, etc.) or indirectly (the Word2 family ALIGN). That is, both constraint families favor alignment, and they work against a single constraint that applies in all cases.

Table 11<INSERT Table 11 HERE>

We illustrate how these constraints work with two tableaux that abstract away from variation: one for a regular, non-alignant word (*hiver*), and one for an alignant word (*hibou*).

(19) Illustration of analysis for *l'hiver* ‘the winter’ and *le hibou* ‘the owl’

<INSERT Tableau 19 HERE>

3.7 MAXIMUM ENTROPY MODEL AND NOISY HARMONIC GRAMMAR MODEL. As in Tagalog, the two models based on constraint weighting produce very similar results. Weights are shown in Table 12.

Table 12<INSERT Table 12 HERE>

Both grammars do a good job of capturing the sigmoid shape of the data (and produce nearly identical predictions), with rates converging at 0 and 1 for the extreme Word2 groups, and maximal Word1 differences for the medial Word2 group, as shown in Figure 18 and Figure 19.

Figure 18<INSERT Figure 18 HERE>

Figure 19<INSERT Figure 19 HERE>

3.8 STOCHASTIC OT MODEL. As in Tagalog, to give the GLA a better chance of finding the best StOT grammar, we ran it 999 times with the Magri update rule and 999 times with the standard, symmetrical update rule, and chose from that set the grammar with the best log likelihood. The best model overall here was one fitted with the standard update rule; its ranking values are listed in Table 13. The large differences in ranking values (e.g. -344.23 vs. -34.52, for the two lowest-ranked constraint, or 90.67 vs. the next value up, 181.11) indicate a failure to converge.

Table 13<INSERT Table 13 HERE>

As shown in Figure 20, the model does a poor job of capturing Word1 differences. For Word1 groups 1-2, the ALIGN constraints practically always outrank NOHIATUS and the USE constraints. For group 3, there is a real competition between ALIGN—group3 (ranking value 188) and NOHIATUS (187), leading to a substantial rate of non-alignant behavior, but none of the USE constraints has a high enough ranking value to compete seriously with ALIGN—group3, so there are no Word1 differences. In group 5, ALIGN—group5 is out of the picture (-344), so the decision depends on the ranking of NOHIATUS versus the USE constraints; the USE constraints from group C provide NOHIATUS with some competition, but there is hardly any difference between the A and B groups (which is why only two traces are clearly visible; A's and B's traces are nearly superimposed), because all of their USE constraints are ranked too low to compete with NOHIATUS.

Figure 20<INSERT Figure 20 HERE>

WHY DOES STOCHASTIC OT FAIL FOR FRENCH? In Tagalog, we argued that Stochastic OT fails because of the phonologically-motivated internal structure of the constraint set. In French, the problem is much more fundamental: the Word1 and Word2 constraint families are SYNERGISTIC, both favoring the same candidate (e.g. *la hache*), because it preserves both the citation form of Word1 and syllable alignment for Word2. The only constraint favoring the other candidate (*l'hache*) is NOHIATUS.

Our straw-man grammar for Tagalog in Table 3, where each consonant had its own constraint, was ANTAGONISTIC: the consonants' constraints favored nasal substitution and the prefixes' constraints penalized it. (Our actual grammar for Tagalog is nearly synergistic: the prefix constraints all penalize nasal substitution, and so do most of the consonant constraints. But one member of the consonant family, *NÇ, favors substitution.)

In unpublished work, Giorgio Magri (p.c.) has proven a property of all grammars of this type, that is, with two synergistic constraint families that are simple (in the sense that for each input, exactly one constraint from each family is applicable) and one opposing constraint. He shows that in such cases, the frequency patterns generated in Stochastic OT will be uniformly converging in one direction and diverging in the other. We saw this for French in Figure 20,

where Word1 differences, though always small, grow monotonically towards the right of the plot, instead of being largest for the middle alignment groups.

To understand this result, we can think of generation in Stochastic OT as a race between constraints that prefer each candidate. In the French case, there are three runners: the anti-alignment team consists of just NOHIATUS, who runs in every trial, and the pro-alignment team sends one ALIGN constraint and one USE constraint. Whoever crosses the finish line first wins the race for their team. If the ALIGN constraint is a fast runner and will probably finish first (as on the left side of Figure 20, where ALIGN—group 1 prevents non-alignment), then it matters little how fast its USE teammate is. The slower ALIGN is, the more it matters which USE constraint is running, because USE bears more responsibility for winning the race. Crucially, this pattern is monotonic: as we move rightwards along the alignment axis in figure Figure 20, to weaker ALIGN constraints, all Word1 differences become (asymptotically) greater.

By contrast, Harmonic Grammar is like a tug-of-war: NOHIATUS pulls the rope in one direction, and an ALIGN constraint and a USE constraint team up to pull it in the other. If ALIGN is strong enough (like ALIGN—group 1) to reliably overpower NOHIATUS on its own, then it matters little which USE constraint is helping it—the help is not needed. If ALIGN is similar in strength to NOHIATUS (ALIGN—group 3), then even small differences between weak USE constraints greatly affect the outcome. But unlike in Stochastic OT, for the weakest ALIGN constraints (ALIGN—group 5), USE differences can decrease again. This will happen when the USE constraints being compared are both substantially weaker than NOHIATUS: without a good boost from ALIGN to move the ALIGN/USE team into competitive territory, the team loses to NOHIATUS nearly all the time anyway. The sigmoid curve families that we saw in Figure 18 and Figure 19—which successfully match the data—are the result.

3.9 ANTTILIAN PARTIAL ORDERING MODEL. As we did for Tagalog, we used the multiple-run technique to identify the best out of 999 Partially Ordered (stratified) OT grammars, by training the GLA but forcing ranking values to remain at least 20 units apart. Because the set of possible stratified grammars is a strict subset of the set of possible StOT grammars, we expect the best

stratified model to be a poorer fit than the best StOT grammar, but we can still assess its qualitative fit to the data.

The best stratified model is shown in Table 14. Although the lower strata appear puzzling—the Word1s are out of order—this is because the rankings of the strata in which the USE constraints appear never matter. For Word2s of groups 1 and 2, the decision to be alignant is made by the first stratum (0% non-alignant behavior). For groups 3 and 4, the decision is made in the second stratum by the ranking of ALIGN and NOHIATUS (50% non-alignant). For group 5, the decision is made in whatever stratum the relevant USE constraint occurs (100% non-alignant)—that is, the USE constraints could all be placed in the same stratum, and there would be no change in the grammar’s behavior. There is thus no opportunity for Word1 differences to manifest.

Table 14<INSERT Table 14 HERE>

The plot of model predictions, in Figure 21, illustrates this lack of Word1 effects; only one trace is visible, because all Word1s behave the same.

Figure 21<INSERT Figure 21 HERE>

3.10 DECISION TREE MODEL. A decision-tree model is conceptually attractive here, as it was in Tagalog: let the morphosyntax give each Word1 a probability of creating either a loose or a close syntactic or prosodic configuration: $[\text{Word1}]_{\text{domain}} [\text{Word2}]_{\text{domain}}$ (loose) or $[\text{Word1 Word2}]_{\text{domain}}$ (close). The loose structure does not allow liaison or élision—alignment is always respected. In the close structure, non-alignment is possible. Let the phonology give each Word2 a probability of triggering liaison or élision if in the close structure. The probability of a Word1+Word2 combination behaving as non-alignant is then the product of two probabilities: the probability that Word1 triggers a close structure and the probability that Word2 triggers liaison or élision.

A decision-tree model was fitted using the `optim()` function in R, maximizing likelihood. The fitted component probabilities are in Table 15, and the result is plotted in Figure 22. For the same reasons as in Tagalog, a decision tree fares poorly: it can capture the ‘pinch’ at one end of

the scale—in this case, a very low non-alignancy probability for Group 1, as desired—but it predicts that rates will spread out (and only slightly) at the other end (Group 5), rather than in the middle. It is possible to create a decision tree in which the Word1s spread out more towards the right of the plot (though the resulting likelihood is lower), by giving the Word1s a greater variety of close-structure probabilities. But it is not possible to achieve the greatest Word1 differentiation at intermediate values of Word2. Differences between Word1s are best preserved when Word2 has the non-alignment probability, 1—that is, at the right end of the plot.

Table 15<INSERT Table 15 HERE>

Figure 22<INSERT Figure 22 HERE>

3.11 COMPARISON OF MODELS. In Table 16, we compare the log likelihoods of the models. MaxEnt and Noisy HG achieve similarly good fits, and the rest do less well. As a baseline, we again include the log likelihood for a hypothetical model that achieves perfect frequency matching for each combination of Word1 and Word2-group.

Table 16<INSERT Table 16 HERE>

3.12 FRENCH CONCLUSIONS. French involves an interaction between a lexical propensity (Word2's propensity to behave as alignant) and unknown morphosyntactic factors or lexical propensities (Word1's propensity to behave as though Word2 is alignant). These two dimensions interact in a pattern correctly modeled by Harmonic Grammar, as either MaxEnt or Noisy HG: at high and low levels of Word2's scale, rates stick to the floor or ceiling; at intermediate levels of Word2's scale, Word1 has its greatest effects. A decision-tree model fails to capture this basic pattern, as do models employing probability distributions over strict-ranking OT (Stochastic OT and Stratified Partial Ordering OT) We showed that the failure of Stochastic OT is a general

result for any analysis in which the two intersecting constraint families (here ALIGN constraints and USE constraints) favor the same candidate.

4. HUNGARIAN VOWEL HARMONY: PHONOLOGY VS. PHONOLOGY. In our third example, Hungarian vowel harmony, we introduce two further elements. First, the intersecting constraint families are both phonological: one family concerns the stem's vowel pattern, and the other concerns its final consonants. Second, for Hungarian we have not just corpus data, but experimental data to support the psychological reality of the patterns observed in the lexicon.

4.1 THE DATA: VOWEL AND CONSONANT EFFECTS AND THEIR INTERSECTION. We have written on Hungarian vowel harmony before (Hayes et al. 2009) and will rely heavily on that material, as well as on the extensive earlier analytic literature cited therein. As in that article, we rely on the dative suffix, which has allomorphs back –nak ([nək]) and front –nek ([nek]).

Backness harmony in Hungarian is largely predictable according to whether the stem's vowels are back (B), front rounded (F), or front unrounded (N, for neutral). Stems whose last vowel is B take back suffixes ([bi:ro:-nək] 'judge-DAT'); stems whose last vowel is F take front suffixes ([sɛmølt-nek] 'wart-DAT'); and some other stem types¹⁹ also invariably take front suffixes. Stems that end in a sequence BN or BNN show variation driven by the conflict between the local, phonetically front trigger N and the distal back trigger B ([høvɛr-nək] 'pal-DAT', [ko:dɛks-nek] 'codex-DAT', [bohe:m-nək] ~ [bohe:m-nek] 'easy-going-DAT'). There is also a fair amount of variation in monosyllabic stems with a neutral vowel (N) ([tsi:m-nek] 'address-DAT', [hɪ:d-nək] 'bridge-DAT').

Among the stem types that vary, we observe two vowel-driven tendencies: the Height Effect (Hayes et al. 2009: 831), where stems ending in [ɛ] take front suffixes more often than stems ending in [e:], which take front suffixes more often than stems ending in [i] or [i:]; and the Count Effect (Hayes et al. 2009: 830), where BNN stems, having two Ns as front triggers, take front suffixes more often than BN stems.

Surprisingly, there are also tendencies driven by the stem's final consonants (Hayes et al. 2009: 836). There are four factors that favor front harmony, given in 20.

- (20) Consonant environments favoring front-voweled suffixes in Hungarian
- a. final bilabial noncontinuant ([p, b, m])
 - b. final sibilant ([s, z, ʃ, ʒ, ts, tʃ, dʒ])
 - c. final coronal sonorant ([n, ɲ, l, r])
 - d. final two-consonant cluster

There is little reason to think that these patterns have any sensible basis in the theory of markedness (phonetic or otherwise), but they do hold true at a statistically significant level in the Hungarian lexicon (Hayes et al. 2009: 839).

To examine the interaction of these two dimensions—vowel pattern and final consonants—we sort the data in a way not employed in Hayes et al. 2009, according to the number of final-consonant factors present: zero (fails to match any of the environments of 20); one (matches one of *a*, *b*, *c*, or *d* in 20), and two (*d* plus one of *a*, *b*, or *c*).²⁰

In Figure 23, we plot the rate, averaged over stems, at which stems take the back suffix allomorph for each vowel pattern and each number of final-consonant properties present, using the same web data as in Hayes et al. 2009. (The data come from 8,915 stems.) The consonant factors show consistent effects within each vowel category: the highest backness rates are found when zero consonant factors are present, and the lowest when two consonant factors are present. We also see the familiar floor and ceiling effects, with backness rates converging to 1 for B stems and to 0 for F stems.

Figure 23<INSERT Figure 23 HERE>

Hayes et al. 2009 used a survey to test nonce BN, BNN, and N stems, and the results, plotted in Figure 24, support the productivity of both the vowel and the consonant effects, and show rates starting to converge to zero for N. (The data come from 1,703 trials and 1,602 distinct stems). The experiment was designed to test vowel patterns expected to show appreciable variation, with no items of the type B, NN, F, and so on. Therefore Figure 24 shows no data for the extremes of the scale. Earlier nonce probe testing (Hayes & Londe 2006: 71) showed that experimental participants essentially treat B as invariantly back and NN and F as invariantly

front; we assume that including these patterns in Figure 24 would complete the familiar sigmoid curve family seen earlier for Tagalog and French. The overall rate of taking the back suffix is shifted lower in the nonce-word data as compared to the lexicon (perhaps because, in the lexicon, lower-frequency words tend to take more front suffixes, and nonce words are the lowest-frequency words of all, Hayes et al. 2009: 830). While in the lexicon backness rates are close to ceiling for Bi words, in the nonce-word data we can observe the effect of final-consonant properties among Bi words. Conversely, in the lexicon final-consonant properties have a noticeable effect within N words, but in the nonce words rates are already close to the floor. Overall, the nonce-word data show that the effects of stem vowel and final consonant carry over to novel items.

Figure 24<INSERT Figure 24 HERE>

4.2 HUNGARIAN MODELING: CONSTRAINTS EMPLOYED. The constraints used in the models to follow are taken directly from Hayes et al. 2009 and shown in Table 17; for detailed discussion of the constraints and their rationale, see Hayes et al. 2009: 834, 836.

Table 17<INSERT Table 17 HERE>

There is some overlapping structure within each family. Within the consonant family, it is possible for both CLUSTER and one of the other three constraints to apply. Within the vowel family, more than one AGREE constraint can apply, especially when the stem ends in a front, unrounded vowel. When the MONOSYLL-[i:] constraint (which is a member of the vowel family because it depends on stem vowels) applies, AGREE(front, local) will also apply. In terms of the synergistic/antagonistic distinction introduced in section 3.8, Hungarian is a mixed case. All of the constraints in the consonant family favor the front suffix, and so do most of the constraints in the vowel family, but two of the vowel constraints favor the back suffix.

4.3 HARMONIC GRAMMAR MODELS: MAXENT AND NHG. Once again, the two constraint-weighting models produce very similar results. Weights are shown in Table 18, and model predictions for lexical items and nonce probes are shown in Figure 25 and Figure 26. Both models capture the basic pattern: rates of taking –nak converge at 100% for B stems, and at 0% for NN and ‘other’ (NNN, BNNN, F, FN, FNN, FNNN), while in the intermediate categories, the consonantal properties show their influence.

The model predictions plotted are for the actual lexicon or wug stems found in each category, allowing comparison to the observed data plotted above. For example, the Bi stems in the lexicon with two final-consonant properties bear a mix of consonantal properties: they all end in a consonant cluster, but some also end in a bilabial, others in a sibilant, and others in a coronal sonorant. This mix is a little different for the Bi stems in the lexicon than for the Bi wug stems, and thus the model’s predictions averaged over those two mixes are a little different. (There is no prediction shown for Be lexical data with two consonant properties, because there are no real words in the lexicon of this type.)

Table 18<INSERT Table 18 HERE>

Figure 25<INSERT Figure 25 HERE>

Figure 26<INSERT Figure 26 HERE>

4.4 STOCHASTIC OT MODELS. As in the Tagalog and French cases, we fitted 999 Stochastic OT grammars with the standard update rule, and 999 with the Magri update rule, and chose the grammar that was the best fit to the lexical data that it was trained on (in this case one found with the standard update rule). The ranking values are shown in Table 19. The extremely low ranking of one constraint indicates that the learner failed to converge.

Table 19<INSERT Table 19 HERE>

The grammar’s predictions for the lexical and nonce-probe data are shown in Figure 27. While some aspects of the pattern are captured—overall higher backness rates on the left sides of

the plots; overall higher backness rates when fewer final-consonant properties hold—the grammar incorrectly predicts that final-consonant effects are wiped out completely in the $B\epsilon$ category, and backness rates do not quite fall to 0 for the ‘other’ category.

Figure 27<INSERT Figure 27 HERE>

Why does the Stochastic OT grammar fail? In the $B\epsilon$ category, there is one constraint favoring *-nak* (AGREE(back, nonlocal)), and two or more favoring *-nek* (AGREE(low front, local), AGREE(non-high front, local), and, if applicable, BILABIAL, CLUSTER, SIBILANT, and CORSON). Whichever of these constraints is top-ranked—usually AGREE(low front, local)—determines the outcome on each trial. The strongest consonantal constraint, BILABIAL, will be top-ranked about 5% of time, which at first seems often enough to matter. But, even when BILABIAL is the top-ranked constraint, the second-ranked constraint is nearly always (about 95% of the time) AGREE(low front, local), meaning that it actually doesn’t matter whether the stem ends in a bilabial consonant or not: if it does, BILABIAL will choose *-nek*, and if it doesn’t, AGREE(low front, local) will also choose *-nek*. It is only in the approximately 0.5% of cases that BILABIAL ranks first, followed by AGREE(back, nonlocal), that it actually matters whether the stem ends in a bilabial consonant. The situation is even starker for the other, weaker consonantal constraints. As in Tagalog and French, we have a soft ranking paradox: to assure low rates of *-nak* for $B\epsilon$ overall, the ranking-value gap between AGREE(low front, local) and AGREE(back, nonlocal) has to be fairly large; but for the consonantal constraints to make a difference, that gap would have to narrow.

The StOT grammar’s more minor failing, having appreciably non-zero *-nak* rates for some of the ‘other’ stems, is due to the BNNN stems in this category. Adding *-nak* to these stems results in two violations of AGREE(double front, local). In the Harmonic Grammar models, each violation contributes to the BNNN-*nak*’s harmony score, pushing its probability very close to zero. But in the Stochastic OT models, BNNN behaves no differently from BNN—the ranking of AGREE(double front, local) is important, but not its number of violations (the competition, for both BNN and BNNN, is always between a candidate that fully satisfies AGREE(double front, local) and one that violates it, whether once or twice). This does not reflect any general failing of

StOT, because the problem could easily be addressed with a constraint like AGREE(triple front, local) that is violated by BNNN-*nak* but not BNN-*nak*.

4.5 PARTIAL-RANKING MODEL. As for Tagalog and French, we approximated a partial-ranking model by running the Gradual Learning Algorithm on the lexical data, using a plasticity of 20, so that all rankings are effectively strict. We did this 999 times and chose the grammar that best fit the lexical data it was trained on. The resulting grammar is shown in Table 20.

Table 20<INSERT Table 20 HERE>

As illustrated in Figure 28, the grammar allows little scope for the consonantal properties to affect backness rates. B words correctly have their fate decided in Stratum 2 and always take the back suffix. Bi words' fate is decided in Stratum 3, again always in favor of backness. Be words are subject to 50-50 competition between two constraints in Stratum 3. Bε words are assigned a 0% backness rate in Stratum 2. In none of these cases do the consonantal constraints come into play. BNN words are more complex. If the final vowel is low, they are never back (Stratum 2); if the final vowel is mid, all three constraints of Stratum 3 are relevant, and the back-favoring constraint AGREE(back, nonlocal) has a 1/3 chance of choosing the back suffix; if the final vowel is high then there is a 50-50 competition in Stratum 3. The apparent variety of rates seen in the plot for lexical BNN results from this variety of final N heights, not from consonantal effects; in the wug data, on the other hand, N heights were balanced, so there is no spurious consonantal effect predicted.

N words that are not monosyllables with [i:] have no constraints favoring backness. Those that are monosyllables with [i:] finally show some consonantal effects: if BILABIAL, CLUSTER, or SIBILANT is relevant, backness rate will be zero. If none of these three is relevant but CORSON is, there will be 50-50 competition in Stratum 6. And if none of the consonantal properties hold, backness will be 100%. NN words, and NN and NNNN (grouped under other) have a predicted backness rate of zero. The variety seen within the 'other' category comes from BNNN words, subject to the same constraints as BNN words.

Figure 28<INSERT Figure 28 HERE>

4.6 COMPARISON OF MODELS. We summarize the models' performance quantitatively in Table 21. As before, we used a backoff of 1/100,000 for predicted probabilities of 0 or 1 that were generated by the non-MaxEnt models. As mentioned at the beginning of this section, we do not include a decision-tree model, because all of the constraints in the Hungarian analysis are purely phonological. In the Tagalog and French cases it was plausible for a morphosyntactic component to feed into a phonological component—with probabilistic decisions being made within each component—but here the constraints must all interact within a single component.

Table 21<INSERT Table 21 HERE>

For all models, the fit to the wug data is worse—even more so in terms of average log likelihood per stem. The reason is that log likelihood is most hurt at intermediate rates, and the wug data, by design, are rich in stems with intermediate backness rates.

4.7 HUNGARIAN CONCLUSIONS. The special aspects of the Hungarian study are as follows. First, Hungarian is a case in which the two intersecting constraint families are both phonological, but the overall data pattern is the same: we find that backness rates for consonant category converge at the extremes of the vowel categories. This is best matched with some version of Harmonic Grammar. Second, unlike in the other cases, we have wug-test data to validate the psychological reality of both dimensions of the lexical pattern.

5. CONCLUSION. In this work, we compared competing frameworks of stochastic grammar by examining how they handle the problem of intersecting constraint families. We did this by finding appropriate examples in which the data are analyzable with intersecting constraint families, and large enough corpora are available.

Figure 4, Figure 17, and Figure 23 give the essential pattern of floor and ceiling effects, where differences in one constraint family are seen only at medial values of the other. As we showed, this qualitative pattern is a direct prediction of Harmonic Grammar, in either of its

stochastic versions (MaxEnt or Noisy HG). Floor and ceiling effects result when the influence of one family is so strong that the other family makes no difference. Distinctions made in the medial region result from the addition of constraint weights that is at the core of this approach.

Other theories were far less successful. The decision-tree theory (§2.6, §3.10) derives not parallel sigmoids but ‘claws’ (Figure 14, Figure 22), with pinching at only one end of the range. Stochastic Optimality Theory can partially derive the crucial pattern in one case (Hungarian), but for Tagalog and French it fails (Figure 9, Figure 20). In the case of Tagalog, the failure arises from a language-specific ranking paradox (§2.6). In French, the failure is more general: Magri’s theorem shows that convergence at both ends of one dimension cannot be obtained with simple, synergistic constraint families (§3.8). The version of Partial Ordering examined here offers a small subset of the grammars available under Stochastic OT, and thus suffers from the same limitations. In addition, the coarseness of the model leads it to collapse distinctions more often than the others.

We conclude that if these three case studies are representative, Harmonic Grammar has the empirical advantage over other constraint models of variation. Results for Noisy HG and MaxEnt were too similar for us to say which flavor of probabilistic Harmonic Grammar captures the data better.

We comment briefly on the generality of our claim. An assumption of both Noisy HG and MaxEnt is that ALL grammatical phenomena are to be derived with weighted constraints: the theory is not some kind of patch on a basically non-stochastic theory of grammar; rather, non-gradient behavior is simply the limiting case that arises when unanimous acquisition data result in extreme weights. Likewise, we do not think it likely that weighted constraints govern only some parts of the grammar; the research tradition arising from Kroch’s pioneering work on syntactic change (1989, 2001) suggests that that the sigmoid curves derived by MaxEnt and NHG models are likely to arise in syntax as well.

A subsidiary finding of our research was the existence of LEXICAL PROPENSITY (first introduced in section 3.4): the continuous, not categorical, tendency of a lexical item to behave in a particular phonological way—for example, to require a preceding syllable boundary or not. Lexical propensity was treated in classical generative phonology through binary approximations, using diacritics or abstract representations (ghost consonants, floating autosegments, etc.). Examining large corpora indicates that such treatment is insufficiently nuanced: it predicts all-or-

nothing variation across stems rather than the gradient patterns we found for Tagalog prefixes and French Word2. To better fit the data, we have encoded lexical propensity as a range of lexically specific constraint weights in Harmonic Grammar.

As a direction for future research, we suggest that psycholinguistic testing of the generalizations for their productivity, beyond what we have already done for Hungarian. Our corpora do represent real language use rather than laboratory data, but they are the result of folding together the speech of a great number of individuals.

Appendix: evidence on French from the literature

CLAIMS THAT THERE IS VARIATION

- Encrevé (1988: 199) observes that the lists of alignant words in four prescriptive sources ‘are far from agreeing—and far from agreeing with what one hears’.²¹
- Dubroca's (1824: 150) pronunciation manual justifies a supplemental chapter on *h*-initial words as follows: ‘Errors are so frequent and so crude in this respect; there are so few men who have sure principles on it’.
- Fouché (1956: 258-265), in a prescriptive manual, notes variation (‘*flottement*’) for several categories of proper names and loans, as well as *hyène* ‘hyena’.
- Grevisse and Goosse (2011: 56), in a widely used prescriptive manual, note that ‘popular speech barely respects disjunction [i.e. lack of liaison or elision] before *h aspiré*’ and that ‘lapses have slipped by authors’. They note cases of:
 - prescriptively non-alignant words that are often treated as alignant (e.g. *hameçon* ‘hook’), and vice versa (e.g. *handicap*);
 - ‘hesitation’ and ‘confusion’, especially in foreign names.
 - published contradictions, including a dictionary that lists *holisme* ‘holism’ as non-alignant, but uses *du holisme* in that very entry.
- Lively discussions can be found on French-language internet forums as to whether a given word is alignant (e.g. *Haitien* ‘Haitian’ and *haricot* ‘bean’).

CLAIMS THAT THERE IS INTRA-SPEAKER VARIATION

- Cohen (1963: 140) cites a government flyer that includes, just two paragraphs apart, both *de un mois* and *d’un mois* ‘of one month’.
- Cohen (1963: 140) also cites a well-regarded history manual that varies between *de Henri IV* and *d’Henri IV* ‘of Henry IV’.
- Grevisse and Goosse (2011: 57-63) note several examples of variation within an individual author.
- Cornulier (1981: 203) claims that ‘a sentence containing both *de Hugo* and *d’Hugo* [‘of Hugo’] would not be remarkable in conversation’.

- Cornulier (1981: 203) sums up the situation thus: ‘many people supposed to speak well seem to flip a coin each time’ when it comes to producing liaison or not in *un handicap* ‘a handicap’, *un HLM* ‘a social housing project’, or *un hameçon* ‘a hook’, ‘but, when questioned, decide in a definitive and sincere fashion, like the grammarians’.
- Gabriel and Meisenburg (2009) found intra-speaker variation in a reading-aloud task.

CLAIMS OF WORD1 DIFFERENCES

Interjections, meta-linguistic uses, and other special words

- Fouché (1956: 137) and Grevisse & Goosse (2011: 61): *oui* ‘yes’ usually behaves as alignant, but *que* sometimes undergoes elision before it.
- Ehrhard (1965: 22): *le oui, les oui, mais oui* (no elision or liaison), but *je crois qu’oui* ‘I think so’ in speech (though generally written *je crois que oui*).
- Grevisse & Goosse (2011: 63): *Onze* ‘eleven’ normally is alignant (takes *le, du*), but can behave as non-alignant with specific Word1s—liaison of *est* ‘is’ is possible in *il est onze heures* ‘it is eleven o’clock’, and *que* and *de* can undergo elision before *onze*.
- Grevisse & Goosse (2011: 65): Names of letters are more likely to behave as non-alignant with Word1s that potentially undergo liaison than with Word1s that potentially undergo elision.

Proper names

- Fouché (1956: 260): Personal names that normally are alignant nevertheless trigger liaison of preceding *ces, des, les, chez*.
- Grevisse & Goosse (2011: 66-67): There is a tendency for any personal name to behave as alignant (even those spelled without <h>, such as *Octave*), the more so if Word1 is *que*, among other factors.
- Malmberg (1943: 44): *Hitler*, as it was becoming more common, was triggering more elision, especially with *de*
- Grevisse & Goosse (2011: 57, seeming to contradict Malmberg, though writing in a different period): ‘We have noted *de Hitler* more often than *d’Hitler*, but the opposite with *que*.’ (though see Malmberg 1943: 44).

- Tranel (1981: 201, fn. 4): Liaison is marginal in *le petit Hollandais*, worse in *les petits Hollandais*, and rejected in *des Hollandais*.
- Durand (1986: 201): Durand's parents say [lə] *Hollandais* 'the Dutch person' (no elision); [le(z)] *Hollandais* 'the Dutch people' (variable liaison); and [pəti.t] *Hollandaise* 'little Dutch woman' (enchaînement).
- Durand (1986: 201): 4 of 10 participants in a reading task said [lə] *hongrois* 'the Hungarian language' (no elision) but e[n] *hongrois* 'in Hungarian' (liaison).

Ordinary words

- Cohen (1963: 140): Cohen's own speech has no elision in *la ouate* 'the cotton wool', but elision in *d'ouate*. Cohen (1972: 103): The same people who say *la ouate* (which Cohen deems more frequent than *l'ouate*) say *un tampon d'ouate* 'a cotton ball'
- Ehrhard (1965: 22): hesitation between *la ouate* and *l'ouate*, but *un tampon d'ouate* is preferred.
- Grevisse & Goosse (2011: 61): '*L'ouate*, not long ago considered outdated [...], is today more common [than *la ouate*], but *la ouate* and especially *de ouate* are still said and written.'
- Tranel (1981: 201, fn. 4): himself accepts liaison in *ils sont hors jeu* 'they are off-side' but not *les hors jeux* 'off-sides' (though there is a part-of-speech difference here).
- Cohen (1963: 140): Cohen's own speech has no liaison in *les hyènes* 'the hyenas' and no elision in *la hyène*, but elision in *d'hyènes*.

LEXICALIZATION OF PARTICULAR WORD COMBINATIONS

- Grevisse and Goosse (2011: 63): *onze* 'eleven' normally behaves as alignant, but triggers elision in the expressions *belle/dame-d'onze-heures* 'grass lily' and, variably, *bouillon-d'/de-onze-heures* 'poisoned beverage'.
- Grevisse and Goosse (2011: 63): *un* 'one' behaves as unaspirated just in *ne faire qu'un* 'to be as one' and *c'est tout un* (with liaison of *tout*) 'it's all the same'.
- Many authors (e.g. L'Huillier 1999: 37): *huit* 'eight' behaves as unaspirated only within certain compound numerals (e.g. *dix-huit* 'eighteen' (liaison), *vingt-huit* (liaison) 'twenty-eight', *trente-huit* 'thirty-eight' (schwa deletion); but as aspirated, with no liaison, in *cent huit* 'one hundred and eight').

- Dubroca (1824: 158): *Hollande* ‘Holland’ is normally aspirated, but behaves as unaspirated in the expressions *toile d’Hollande* ‘type of cloth’ and *fromage d’Hollande* ‘type of cheese’
- Dubroca (1824: 159): *Hongrie* ‘Hungary’ is unaspirated just in *point d’Hongrie* ‘type of parquetry’ and *eau de la reine d’Hongrie* ‘type of perfume’, though we find that texts show variation for all of these.
- Cohen (1963: 138): *hameçon* ‘hook’ is normatively aspirated and greatly varying in speech, but is always unaspirated in *mordre à l’hameçon* ‘to bite the hook’ .

THE WORD *HIATUS*

Prescriptive sources used to list this word as unaspirated, but many now give it as aspirated.

- Grevisse and Goosse (2011: 60) note that the dictionary of the Académie française made the switch in its 1997 edition.

SCHWA/ZERO ALTERNATIONS

- Grammont (1914: 124): schwa in alignant [*ynə*] *hache* ‘an axe’, vs. no schwa in consonantal [*yn*] *tache* ‘a spot’ and non-alignant [*yn*] *honnête femme* ‘an honest woman’
- Prescriptive sources and analyses of normative French find this schwa largely when the preceding word is spelled with final *e*, and analyze the phenomenon as blocking of deletion: Schane (1968), Selkirk (1972: 329, 379), Dell (1973: 84-93), Fouché (1956: 135-139), Grevisse & Goosse (2011: 55, 63).
- Tranel (1981: 287) has observed schwa in *quel hasard* ‘such a coincidence’, has obtained schwa in direct elicitation of *sept haies* ‘seven hedges’, and cites Martinon (1913: 249) for schwa in *avoir honte* ‘to be ashamed’.
- Pagliano (2003) notes that schwa is possible for *quel hublot* ‘which porthole’.
- Cohen (1963: 137) reports that his own speech has schwa in *une hâte* ‘a haste’, *une haie* ‘a hedge’, but not in *une hallebarde* ‘a halberd’ (all *h-aspiré* words).

PRE-GENERATIVE ALIGNMENT ANALYSES

- Frei (1929: 96-100), in a study of errors and non-normative pronunciations, puts forth the general principle that clarity is enhanced when elements of an utterance are sufficiently separated. In particular, a word boundary should coincide with a syllable boundary—this is

frequently untrue in French, because of rampant resyllabification across syllable boundaries (*enchaînement*), but nonetheless, Frei proposes, French speakers are subject to this drive. He writes, ‘such is the real *raison d’être* of the so-called *h aspiré*, which is in reality a separator designated to make the syllable boundary correspond with the word boundary; popular language has a tendency to extend it: un | *huissier*’ (i.e. without liaison, although *huissier* ‘bailiff’ is, normatively, non-alignant) (p. 97). Frei relates the lack of liaison and elision, and the insertion/non-deletion of preceding schwa seen in alignant words to similar ‘errors’ seen with ordinary words, as in the popular lack of elision in *quelque chose de ample* ‘something full’ (p. 97).

- Damourette and Pichon (1930: 198) characterized *h-aspiré* as an ‘assurance of hiatus’.
- Malmberg (1943: 35-36) writes of ‘a tendency to isolate the word from its context and not let it be part of a phonetic group, to better conserve its individuality’.

OBSERVATIONS ON ALIGNMENT AS A WAY TO SET APART CERTAIN WORDS

Metalinguistic uses

- Grevisse & Goosse (2011: 64): Any vowel-initial word can be treated as alignant when used metalinguistically, as in *diminutif de aube* ‘diminutive of dawn’.
- Fouché (1956): *de* is normally used before citing a word when one wishes to draw attention to its initial vowel (*l’a de « avoir »* ‘the *a* of *to have*’), but there is variation when one wants to draw attention to a later sound (*le double f d(e) « effet »* ‘the double *f* of *effect*’).
 - Our interpretation is that *avoir* has even stronger reasons to begin with a syllable boundary than *effet*, because not only the word as a whole but also the initial sound specifically is being set apart.

Foreign, archaic, or technical character:

- Martinet (1933: 201-202): many alignant words are archaic or technical.
- Grevisse and Goosse (2011: 66-67): alignancy of a personal name is more likely when the name sounds particularly foreign.
- Grevisse and Goosse (2011: 67): there is a tendency towards alignancy when a word or phrase is felt to be foreign, even if not spelled with <h> (*la upa-upa* ‘the upa-upa [Tahitian dance]’).

A foreign word might be expected to lose its alignant character as it becomes more familiar:

- Ehrhard (1965: 23): Sartre wrote alignant *le ukulélé* ‘the ukulele’ in 1945, but twenty years later, with the word having become common, non-alignant *l’ukulélé* was the norm.²²

Length, alone and in interaction with metalinguistic uses or foreign/archaic/technical character

- Grevisse and Goosse (2011: 56): popular speech often fails to respect a word’s status as alignant, but ‘the shorter the word, the better disjunction is maintained, as liaison and élision would render it difficult to identify’.
- Grevisse and Goosse (2011: 66-67): disjunction before a personal name is more likely when the name is short, and especially if it is only an initial (e.g. *A Meillet*).
- Grevisse and Goosse (2011: 64): while alignancy is generally optional in metalinguistic uses, it is nearly obligatory when the item is a monosyllabic function word or affix.
- Grevisse and Goosse (2011: 65): similarly, when another part of speech is used as a noun, it often becomes alignant if monosyllabic (*le on* ‘the one/we’).
- Fouché (1956: 264-265): foreign-origin names of French persons are alignant if monosyllabic (e.g. *Hahn*), but tend to be non-alignant if polysyllabic. The same applies to German geographic names in France and is the tendency for indigenous names in French overseas territory.
- Malmberg (1943: 44) conjectures, following Gougenheim (1938), that the real reason behind instances where a base word is alignant and its derived forms non-alignant is length: alignant *Hanse* ‘Hanseatic League’, *hile* ‘hilum’, *héros* ‘hero’, *Hitler*, *Hégel* are shorter than their non-alignant or varying derivatives *hanséatique* ‘Hanseatic’, *hilifère* ‘hiliferous’, *héroïne* ‘heroine’, *héroïsme* ‘heroism’, *hitlérien* ‘Hitlerian’, *hégélianisme* ‘Hegelianism’ (see also Fouché 1956: 258, 263).
- Plénat (1995: 7): words formed by the language disguise Verlan are alignant if monosyllabic, such as *ouf* from *fou* ‘crazy’.
- Dell (1970: 86-90) and Tranel (1981): schwa insertion is more common for alignant words that are monosyllabic than for longer alignant words (Dell’s analysis is in terms of stress).

REFERENCES

- ANTTILA, ARTO. 1997a. Deriving variation from grammar. *Variation, change, and phonological theory*, ed. by Frans Hinskens; Roeland van Hout and W. Leo Wetzels, 35–68. Amsterdam: John Benjamins.
- ANTTILA, ARTO. 1997b. Variation in Finnish phonology and morphology. Stanford, CA: Stanford dissertation.
- BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER and STEVEN WALKER. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1–6. Online: <http://CRAN.R-project.org/package=lme4>.
- BERGER, ADAM L.; STEPHEN A. DELLA PIETRA and VINCENT J. DELLA PIETRA. 1996. A maximum entropy approach to natural language processing. *Computational linguistics* 22.39–71.
- BERKO, JEAN. 1958. The child's learning of English morphology. *Word* 14.150–177.
- BLUST, ROBERT. 2004. Austronesian nasal substitution: a survey. *Oceanic linguistics* 43.73–148.
- BOERSMA, PAUL. 1998a. *Functional phonology: formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- BOERSMA, PAUL. 1998b. Spreading in functional phonology. *Proceedings of the International Congress of Phonetic Sciences* 22.1–20.
- BOERSMA, PAUL. 2001. Review of Arto Anttila: *Variation in Finnish phonology and morphology*. *GLOT international* 5.33–40.
- BOERSMA, PAUL. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117.1989–2054.
- BOERSMA, PAUL and Joe PATER. 2013. Convergence properties of a gradual learning algorithm for harmonic grammar. University of Amsterdam and University of Massachusetts, Amherst, MS.
- BREIMAN, LEO; JEROME FRIEDMAN; CHARLES STONE and RICHARD A. OLSHEN. 1984. *Classification and regression trees*. 1st edition. Boca Raton: Chapman and Hall/CRC.
- CHOMSKY, NOAM. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- CLEMENTS, GEORGE. N and SAMUEL JAY KEYSER. 1983. *CV phonology: a generative theory of the syllable*. Cambridge, MA: MIT Press.

- COHEN, MARCEL. 1963. *Nouveaux regards sur la langue française*. Paris: Editions Sociales.
- COHEN, MARCEL. 1972. *Une fois de plus des regards sur la langue française*. Paris: Éditions Sociales.
- CORNULIER, BENOÎT DE. 1981. H-aspirée et la syllabation: expressions disjonctives. *Phonology in the 1980s*, ed. by Didier L. Goyvaerts, 183–230. Ghent: Story-Scientia.
- DAMOURETTE, JACQUES and ÉDOUARD PICHON. 1930. *Des mots à la pensée: essai de grammaire de la langue française*. Vol. 1. Paris: Éditions d'Artrey.
- DELL, FRANÇOIS. 1970. Les règles phonologiques tardives et la morphologie dérivationnelle du français. Cambridge, MA: MIT dissertation.
- DELL, FRANÇOIS. 1973. *Les règles et les sons*. Paris: Hermann.
- DELLA PIETRA, STEPHEN; VINCENT J. DELLA PIETRA and JOHN D. LAFFERTY. 1997. Inducing features of random fields. *IEEE transactions on pattern analysis and machine intelligence* 19.380–393.
- DJALALI, ALEX J. 2013. A constructive solution to the ranking problem in partial order optimality theory. Stanford, CA: Stanford University, MS.
- DUBROCA, LOUIS. 1824. *Traité de la prononciation des consonnes et des voyelles finales des mots français, dans leur rapport avec les consonnes et les voyelles initiales des mots suivans; suivi de la prosodie de la langue française*. Paris: chez l'Auteur, Delaunay, and A. Johanneau.
- DURAND, JACQUES. 1986. French liaison, floating segments and other matters in a dependency framework. *Dependency and non-linear phonology*, ed. by Jacques Durand, 161–201. London: Croom Helm.
- EHRHARD, JEAN. 1965. *Remarques sur trois difficultés de la prononciation française*. Saigon: J. Aspar.
- ENCREVÉ, PIERRE. 1988. *La Liaison avec et sans enchaînement: phonologie tridimensionnelle*. Paris: Seuil.
- ENGLISH, LEO. 1986. *Tagalog-English dictionary*. Manila: Congregation of the Most Holy Redeemer; distributed by Philippine National Book Store.
- FLACK, KATHRYN GILBERT. 2007. The sources of phonological markedness. Amherst: University of Massachusetts dissertation.
- FOUCHÉ, PIERRE. 1956. *Traité de prononciation française*. Paris: C. Klincksieck.

- FREI, HENRI. 1929. *La grammaire des fautes: introduction à la linguistique fonctionnelle, assimilation et différenciation, brièveté et invariabilité, expressivité*. Paris: Librairie Paul Geuthner.
- FRUEHWALD, JOSEF. 2012. Redevelopment of a morphological class. *University of Pennsylvania working papers in linguistics* 18(1). Online:
<http://repository.upenn.edu/pwpl/vol18/iss1/10>.
- GAATONE, DAVID. 1978. Phonologie abstraite et phonologie concrète: A propos de h-aspiré en français. *Lingvisticae investigationes* 2.3–22.
- GABRIEL, CHRISTOPH and TRUDEL MEISENBURG. 2009. Silent onsets? An optimality-theoretic approach to French h aspiré words. *Variation and gradience in phonetics and phonology*, ed. by Caroline Féry; Frank Kügler and Ruben van de Vijver, 163–184. Berlin: Mouton de Gruyter.
- 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, 111–120. Stockholm: Stockholm University.
- GOUGENHEIM, GEORGES. 1938. *Système grammatical de la langue française*. Paris: Bibliothèque du Français moderne, J.L.L. d'Artrey A. Ronteix-d'Artrey.
- GRAMMONT, MAURICE. 1914. *Traité pratique de prononciation française*. Paris: Delagrave.
- GREF, EMILY. 2008. Biphasal and monophasal spellout: evidence from Austronesian nasal substitution. Montreal: McGill University B.A. thesis.
- GREVISSE, MAURICE and ANDRÉ GOOSSE. 2011. *Le bon usage: grammaire française*. 15th ed. Bruxelles: De Boeck-Duculot.
- HAYES, BRUCE and ZSUZSA CZIRÁKY LONDE. 2006. Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony. *Phonology* 23. 59–104.
- HAYES, BRUCE; BRUCE TESAR; and KIE ZURAW. 2014. OTSoft 2.3.3.
<http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- HAYES, BRUCE; KIE ZURAW; ZSUZSA CZIRÁKY LONDE; and PETER SIPTÁR. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822–863.
- JÄGER, GERHARD and ANETTE ROSENBACH. 2006. The winner takes it all – almost: cumulativity in grammatical variation. *Linguistics* 44.937–971.

- JAYNES, EDWIN T. 1957. Information theory and statistical mechanics. *Physical review, series II* 106.620–630.
- KAYE, JONATHAN and JEAN LOWENSTAMM. 1984. De la syllabicité. *Forme sonore du langage: structure des représentations en phonologie*, ed. by François Dell; Daniel Hirst and Jean-Roger Vergnaud, 123–159. Paris: Hermann.
- KROCH, ANTHONY. 2001. Syntactic change. *Handbook of contemporary syntactic theory*, ed. by Mark Baltin and Chris Collins, 699–729. Oxford, UK: Blackwell.
- KROCH, ANTHONY S. 1989. Reflexes of grammar in patterns of language change. *Language variation and change* 1.199–244.
- LEGENDRE, GERALDINE; YOSHIRO MIYATA and PAUL SMOLENSKY. 1990. Harmonic grammar – a formal multi-level connectionist theory of linguistic well-formedness: an application. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* 12.884–891. Mahwah, NJ: Lawrence Erlbaum Associates.
- LEGENDRE, GÉRALDINE; ANTONELLA SORACE and PAUL SMOLENSKY. 2006. The optimality theory–harmonic grammar connection. *The harmonic mind*, ed. by Paul Smolensky and Géraldine Legendre, 339–402. Cambridge, MA: MIT Press.
- L’HUILIER, MONIQUE. 1999. *Advanced French grammar*. Cambridge: Cambridge University Press.
- LIN, YURI; JEAN-BAPTISTE MICHEL; EREZ LIEBERMAN AIDEN; JON ORWANT; WILL BROCKMAN and SLAV PETROV. 2012. Syntactic annotations for the Google Books Ngram Corpus. *Proceedings of the ACL 2012 system demonstrations*, 169–174. (ACL ’12). Stroudsburg, PA: Association for Computational Linguistics.
- MAGRI, GIORGIO. 2012. Convergence of error-driven ranking algorithms. *Phonology* 29.213–269.
- MALMBERG, BERTIL. 1943. *Le système consonantique du français moderne: études de phonétique et de phonologie*. (Études Romanes de Lund VII). Lund: Alf Lombard.
- MARTINET, ANDRÉ. 1933. Remarques sur le système phonologique du français. *Bulletin de la Société de linguistique de Paris* 34.191–202.
- MARTINON, PHILIPPE. 1913. *Comment on prononce le français: traité complet de prononciation pratique avec les noms propres et les mots étrangers*. Larousse.

- MCCARTHY, JOHN J. and ALAN PRINCE. 1993. Generalized alignment. *Yearbook of morphology*. 79–154.
- MCCARTHY, JOHN J. and ALAN PRINCE. 1995. Faithfulness and reduplicative identity. *University of Massachusetts occasional papers in linguistics*, ed. by Laura Walsh Dickey; Jill N. Beckman and Suzanne Urbanczyk, 249–384. Amherst, MA: University of Massachusetts.
- MICHEL, JEAN-BAPTISTE; YUAN KUI SHEN; AVIVA PRESSER AIDEN; ADRIAN VERES; MATTHEW K. GRAY; JOSEPH P. PICKETT; DALE HOIBERG; et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014).176–182.
- NEWMAN, JOHN. 1984. Nasal replacement in western Austronesian : an overview. *Philippine journal of linguistics* 15–16.1–17.
- PAGLIANO, CLAUDINE. 2003. L'épenthèse consonantique en français: ce que la syntaxe, la sémantique et la morphologie peuvent faire à la phonologie : parles-en de ta numérotation ?impossible. Nice: University of Nice Sophia-Antipolis dissertation.
- PATER, JOE. 1999. Austronesian nasal substitution and other NC effects. *The prosody-morphology interface*, ed. by René Kager; Harry van der Hulst and Wim Zonneveld, 310–343. Cambridge: Cambridge University Press.
- PATER, JOE. 2001. Austronesian nasal substitution revisited: What's wrong with *NC̥ (and what's not). *Segmental phonology in optimality theory: constraints and representations*, ed. by Linda Lombardi, 159–182. Cambridge: Cambridge University Press.
- PATER, JOE. 2008. Gradual learning and convergence. *Linguistic inquiry* 39.334–345.
- PETERS, CATHERINE A. 2002. Statistics for analysis of experimental data. *Environmental engineering processes laboratory manual*, ed. by Susan E Powers. Champaign, IL: Association of Environmental Engineering and Science Professors Foundation.
- PLÉNAT, MARC. 1995. Une approche prosodique de la morphologie du verlan. *Lingua* 95.97–129.
- 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.
- PRINCE, ALAN and PAUL SMOLENSKY. 2004. *Optimality theory: constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell.

- R CORE TEAM. 2014. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Online: www.R-project.org.
- SAS INSTITUTE INC. 2014. *SAS software*.
- SCHANE, SANFORD. 1968. *French phonology and morphology*. Cambridge, MA: MIT Press.
- SCHANE, SANFORD. 1978. L'emploi des frontières de mots en français. *Etudes de phonologie française*, ed. by Benoît de Cornulier and François Dell, 133–147. Paris: Editions du CNRS.
- SELKIRK, ELISABETH. 1972. The phrase phonology of English and French. Cambridge, MA: MIT dissertation.
- SELKIRK, ELISABETH and JEAN-ROGER VERGNAUD. 1973. How abstract is French phonology? *Foundations of language* 10.249–254.
- SHIH, STEPHANIE. 2014. Towards optimal rhythm. Stanford, CA.: Stanford dissertation.
- SMOLENSKY, PAUL. 1986. Information processing in dynamical systems: foundations of harmony theory. *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, ed. by David E. Rumelhart, James L. McClelland and PDP Research Group, 194–281. Cambridge, MA: Bradford Books/MIT Press.
- TRANEL, BERNARD. 1981. *Concreteness in generative phonology: evidence from French*. Berkeley: University of California Press.
- TRANEL, BERNARD. 1995. Current issues in French phonology: liaison and position theories. *The handbook of phonological theory*, ed. by John A. Goldsmith, 798–816. Cambridge, MA, and Oxford, UK: Blackwell.
- TRANEL, BERNARD and FRANCESCA DEL GOBBO. 2002. Local conjunction in Italian and French phonology. *Romance phonology and variation: selected papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000*, 191–218. (Current Issues in Linguistic Theory). Amsterdam and Philadelphia: Benjamins.
- WALKER, DOUGLAS C. 2001. *French sound structure*. University of Calgary Press.
- WILSON, COLIN. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive science* 30.945–982.
- WILSON, COLIN. 2014. Maximum entropy models (tutorial presentation). MIT.
- ZURAW, KIE. 2000. Patterned exceptions in phonology. University of California, Los Angeles dissertation.

ZURAW, KIE. 2009. Frequency influences on rule application within and across words.

Proceedings of CLS (Chicago Linguistic Society) 43. 283–309.

ZURAW, KIE. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28. 417–472.

NOTES

¹ We don't require that structure of the constraint set be as simple as in 0: for example, there could be a constraint applying to more than one value of a dimension (such as non-mid vowels).

² There is a minor correction: Zuraw 2010 gives a count of 1 substituting /g/-initial word for *maŋ-RED-*, but the correct count is 0.

³ The reference to entropy is explicated by Jaynes (1957: 620), who stated that 'the maximum-entropy estimate ... is the least biased estimate possible on the given information; i.e. it is maximally noncommittal with regard to missing information.' For example, in an OT-style MaxEnt grammar, a tableau with nothing but zeros for the constraint weights results in equal probability for all candidates—the least committal, most entropic distribution.

⁴ Our notation gives constraints positive weights and negates the weighted sum prior to exponentiation, so that 'Harmony' is in effect a penalty score. This is a notational variant (more intuitive to us, and apparently to its originator, Wilson 2006). The alternative, which is widely used and has historical priority (Smolensky 1986), is to let the constraint weights be negative and avoid the negative sign in the formula for probability; this lets 'Harmony' be related directly to higher probability—as the word connotes—though the Harmony values are themselves negative. Absolutely nothing empirical hinges on this distinction.

⁵ We assume that all unselected candidates are penalized by constraints whose weights are high enough to make their predicted probabilities negligible.

⁶ Violations are often expressed as negative numbers instead. Either way, the best weighted sum is the one closest to zero.

⁷ Default initial weights: all 0. Initial plasticity: 0.01; final plasticity: 0.001. Number of learning trials: 10,000,000. Noise imposed is a Gaussian random variable (mean 0, standard deviation 1), multiplied by 2 (Boersma 1998a). The only bias imposed is that weights may not be negative.

⁸ Because the noise variables are distributed symmetrically around zero, the distribution of *noise+noise-noise-noise* is the same as the distribution of *noise+noise+noise+noise*.

⁹ It was computationally convenient to implement 999 trials rather than 1000. Initial ranking values: 100 for all constraints (i.e. no biases). Number of learning trials: 1,000,000. Starting plasticity: 2; ending plasticity: 0.01. Noise: 2.

¹⁰ We define stratification more explicitly as follows: if constraints B and C are unranked, then every constraint that outranks B must also outrank C, and every constraint that B outranks must also be outranked by C. Then we can say that B and C are in the same ranking stratum, as in $A \gg \{B,C\} \gg D$.

The full version of Anttila's theory is more powerful because it allows a grammar like $A \gg B$, with C unranked. This means that there are three equally probable rankings: $C \gg A \gg B$, $A \gg C \gg B$, $A \gg B \gg C$. This is not possible in Stochastic OT, where if A's ranking value is so much higher than B's that A effectively always outranks B, then there is no possible ranking value for C that makes it likely to outrank A but also to be outranked by B. Since the GLA only learns possible stochastic OT grammars, it follows that such rankings cannot be learned by the GLA.

¹¹ See Djalali (2013) for an algorithm that inputs non-quantitative data (candidates either are legal or not), finds the compatible partial orderings, and calculates predicted frequencies. Limitations in the software implementation available to us have prevented us from testing this algorithm on the three cases examined here.

¹² The Partial Ordering model predicts some 0 rates, whose log can't be taken. We applied a back-off of 1/100,000 (because there were 100,000 test trials), treating 0 as 0.00001 (and the other candidate's 1 as 0.99999). We did the same for the baseline perfect model.

¹³ Some authors reserve the term *liaison* for words with a single spelled form, whose final consonant is pronounced or not according to context, and call the cases in 0 allomorphy.

¹⁴ There can be an overt [h] in some dialectal, theatrical, and emphatic contexts.

¹⁵ Smoothing set to 3, which means that each value plotted is the average of the year shown and three years on either side.

¹⁶ In a similar spirit, Tranel (1981) proposes that emphasis and citation are reasons for a word to be exempt from normal syllabification rules; being lexically alignant means having a lexical requirement to be similarly exempt.

¹⁷ Our analysis below will not explicitly distinguish vowels and glides, though they do have some differences in behavior, not detectable in the written data we use. For example, Tranel (1981: 304) points out that optional schwa deletion occurs before consonants but not alignant vowels in spoken French: *ce vin* [s(ə) vɛ̃] 'this wine', but *ce hasard* [sə azaʁ] 'this coincidence'; but, he notes, optional schwa deletion is allowed before an alignant glide (*ce whisky-là* [s(ə)

wiski la], in violation of ALIGN. Another way in which alignant glides differ from alignant vowels is that in dialects where *cinq* ‘five’ loses the final consonant of its citation form before a consonant-initial word, it also can lose it before an alignant glide, but not an alignant vowel (Cornulier 1981: 206).

¹⁸ In the logistic regression model, the coefficients are: *beau/bel* 2.46, *ma/mon* 2.40, *vieux/vieil* 1.76; *de/d’* 0.31, *au/à l’* 0.00, *la/l’* -0.16; *le/l’* -0.37, *du/de l’* -0.49. For details on exactly which pairwise differences are significant and according to what criteria, see the supplemental files for French.

¹⁹ We will use F both for stems that end in F and for stems that end in F followed by any number of N, because all such stems consistently take front suffixes. NNN, NNNN, and BNNN stems are also consistently front in our data.

²⁰ In principle, there are eight possible consonantal-constraint patterns (viz.: has none of the properties in **Error! Reference source not found.**; has just *a*, just *b*, just *c*, just *d*; has *a* and *d*, *b* and *d*, or *c* and *d*), and it would be ideal to treat each one separately. But 12% of the points in such a plot would be missing because there is no data, and another 28% would be based on fewer than 10 stems. By grouping the consonantal dimension into the three values used here, we end up with just one missing data point in 0 is missing, and only three more (12%) based on fewer than ten stems.

²¹ All translations are our own.

²² Though it is possible that, under English influence, there is variation between [jukulele] and [ukulele]

rates of vowel devoicing	constraint applying to high Vs	constraint applying to mid Vs	constraint applying to low Vs
constraint applying to aspirated Cs	/p ^h i/ → [p ^h _i]: <i>x%</i>	/p ^h e/ → [p ^h _e]: <i>y%</i>	/p ^h a/ → [p ^h _a]: <i>z%</i>
constraint applying to voiceless unaspirated Cs	/pi/ → [p _i]: <i>w%</i>	/pe/ → [p _e]: <i>v%</i>	/pa/ → [p _a]: <i>u%</i>
constraint applying to voiced Cs	/bi/ → [b _i]: <i>t%</i>	/be/ → [b _e]: <i>s%</i>	/ba/ → [b _a]: <i>r%</i>

FIGURE 1. Intersecting families of constraints: a schematic view

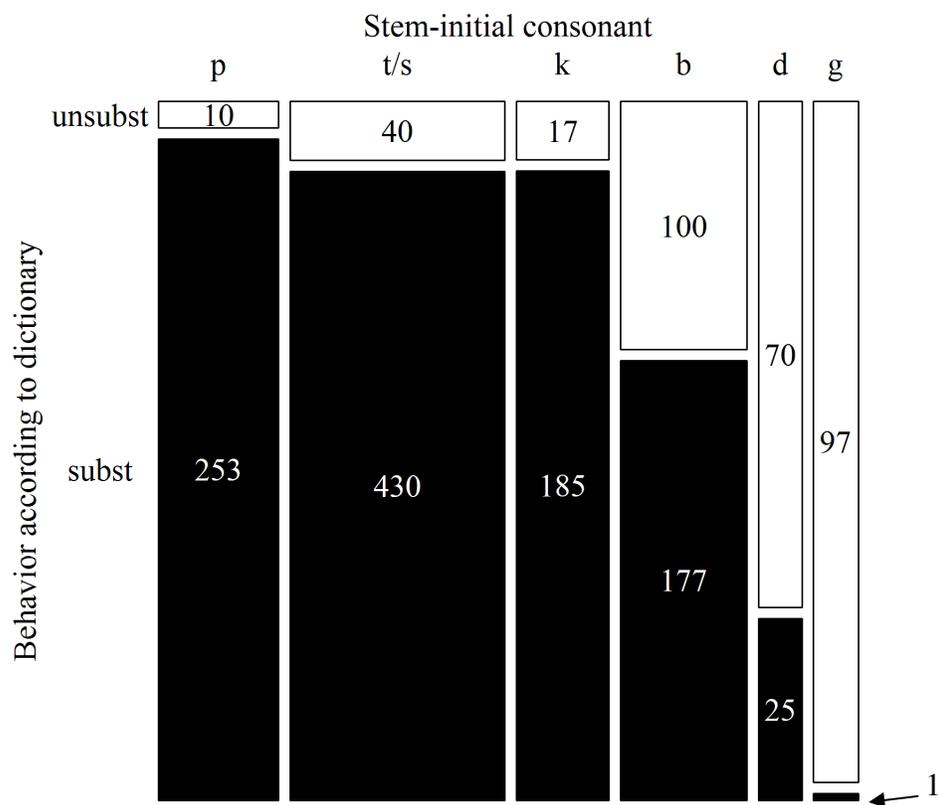
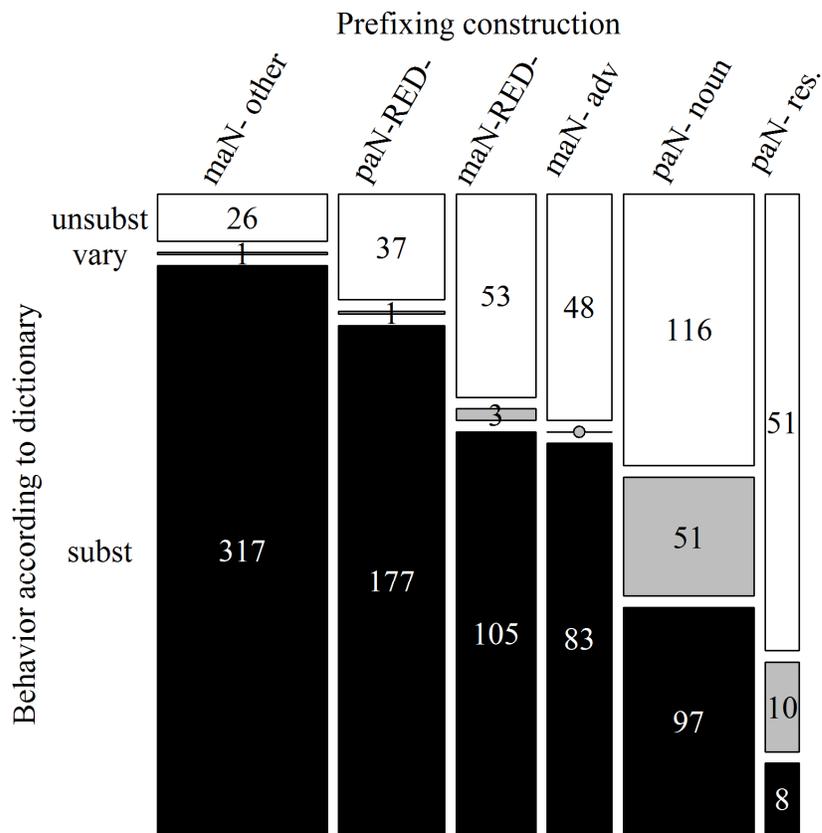


FIGURE 2. Rates of nasal substitution by consonant — dictionary data



Key (RED = reduplication):

<i>Prefix</i>	<i>Construction</i>
a. /maŋ _{-other} /	non-adversative verbs
b. /paŋ _{-RED-} /	mainly gerunds
c. /maŋ _{-RED-} /	professional or habitual nouns
d. /maŋ _{-adv} /	adversative verbs
e. /paŋ _{-noun} /	various nominalizations
f. /paŋ _{-res} /	reservational adjectives

FIGURE 3. Tagalog Nasal Substitution by prefix construction

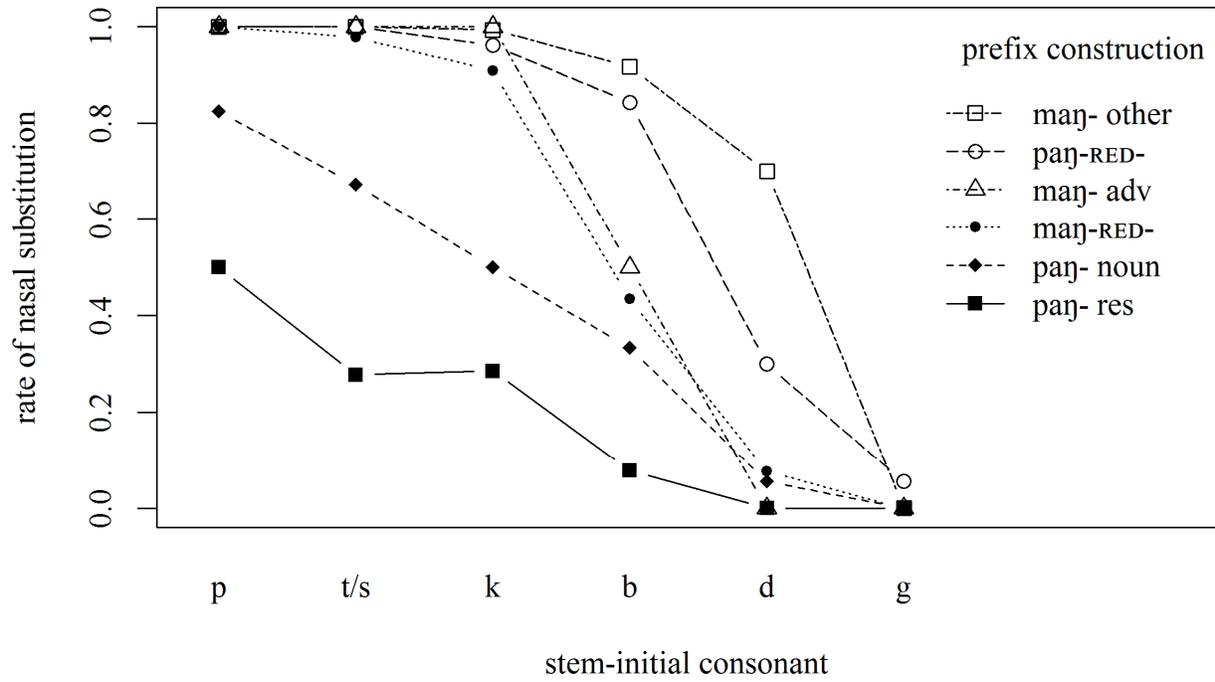


FIGURE 4. Rates of nasal substitution collated

	<i>frequency</i>	NASSUB	*NC _o	*[_{root} m/n/ŋ]	*[_{root} n/ŋ]	*[_{root} ŋ]	UNIF-maŋ-OTHER	UNIF-paŋ-RED-	UNIF-maŋ-RED-	UNIF-maŋ-ADV	UNIF-paŋ-NOUN	UNIF-paŋ-RES
/maŋ _{-other} +p.../												
[ma-m...]	65			*			*					
[mam-p...]	0	*	*									
/paŋ _{-res} +p.../												
[pa-m...]	4.5			*								*
[pam-p...]	4.5	*	*									
/maŋ _{-other} +t.../												
[ma-n...]	71			*	*		*					
[man-t...]	0	*	*									
/maŋ _{-other} +k.../												
[ma-ŋ...]	74.5			*	*	*	*					
[maŋ-k...]	0.5	*	*									
/maŋ _{-other} +b.../												
[ma-m...]	66			*			*					
[mam-b...]	6	*										
/maŋ _{-other} +d.../												
[ma-n...]	7			*	*		*					
[man-d...]	3	*										
/maŋ _{-other} +g.../												
[ma-ŋ...]	0			*	*	*	*					
[maŋ-g...]	13	*										

<Tableau for example 7>

<i>weight:</i>	4.9	0.0	2.1	1.2	2.3	0.0	0.8	1.9	2.3	4.1	6.0			
<i>/paŋ</i> _{Noun+t...} <i>/</i>	*NC _o	* _[root m/n/ŋ]	* _[root n/ŋ]	* _[root ɲ]	NASSUB	UNIF-maŋ-OTHER	UNIF-paŋ-RED-	UNIF-maŋ-ADV	UNIF-maŋ-RED-	UNIF-paŋ-NOUN	UNIF-paŋ-RES	<i>Harmony</i> (10)	<i>eHarmony</i> (11)	<i>probability</i> (8), (12)
[pa-n...] (subst.) 60.5 words (67%)	0	1	1	0	0	0	0	0	0	1	0	1 × 0.0 + 1 × 2.1 + 1 × 4.1 = 6.2	$e^{-6.2}$ = 0.0020	$\frac{.0020}{.0020+.0008} =$ 0.73
[pan-t...] (unsubst.) 29.5 words (33%)	1	0	0	0	1	0	0	0	0	0	0	1 × 4.9 + 1 × 2.3 = 7.2	$e^{-7.2}$ = 0.0008	$\frac{.0008}{.0020+.0008}$ = 0.27

<Tableau for Example 9>

<i>constraint</i>	<i>weight</i>
NASUB	2.31
*NC̣	4.85
*[_{root} m/n/ŋ]	0.00
*[_{root} n/ ŋ]	2.13
*[_{root} ŋ]	1.16
UNIF-man-OTHER	0.00
UNIF-paŋ-RED	0.82
UNIF-man-ADV	1.92
UNIF-man-RED	2.29
UNIF-paŋ-NOUN	4.06
UNIF-paŋ-RES	6.01

TABLE 1. MaxEnt weights fitted to Tagalog

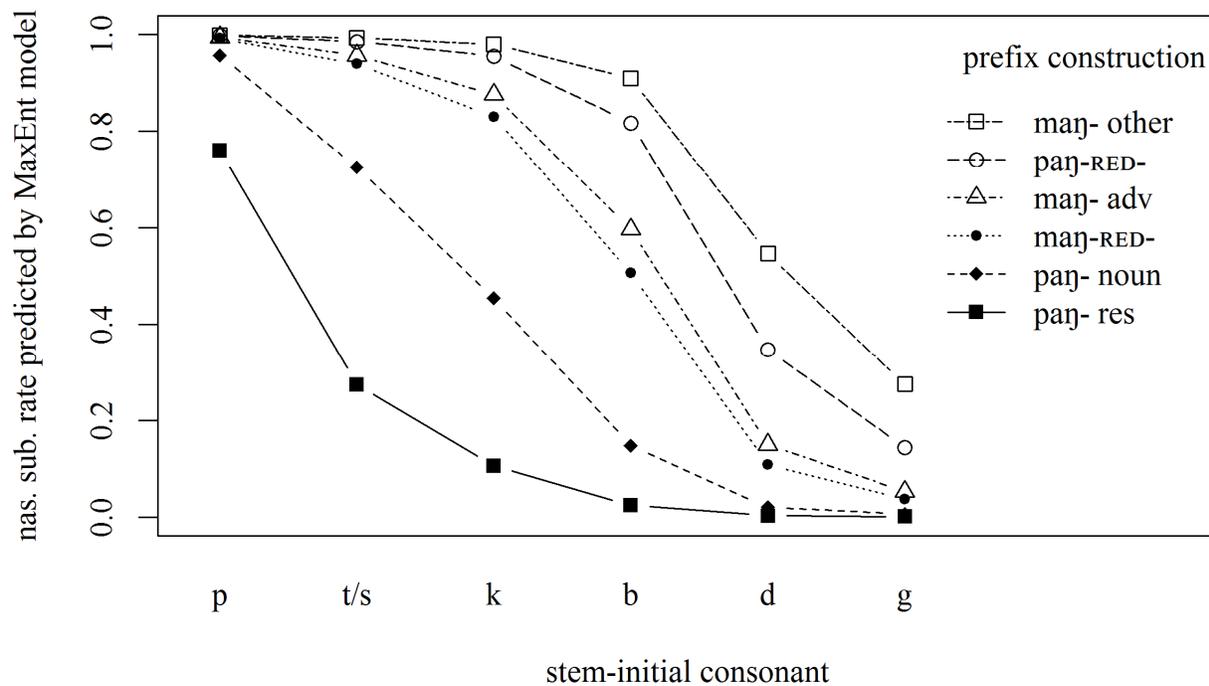


FIGURE 5. MaxEnt grammar's probabilities

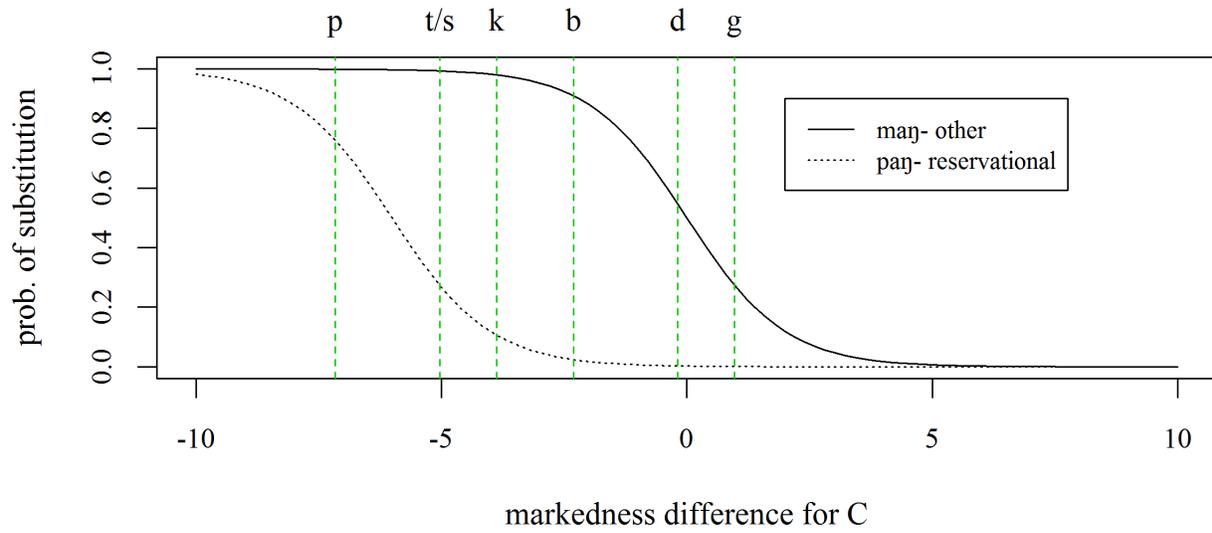


FIGURE 6. Substitution probabilities for two prefixes, as a function of consonants' markedness differences

<i>weight</i>	10.6	0.4	4.4	2.2	6.2	0.4	2.1	4.3	5.0	8.9	13.6		
<i>noise (Gaussian)</i>	-0.7	0.3	1.2	0.1	0.2	0.0	-0.5	-0.2	-0.4	-0.4	-1.2		
<i>weight+noise</i>													
<i>/paŋ-Noun+t.../</i>	*NC _o	*[_{root} m/n/ŋ]	*[_{root} n/ŋ]	*[_{root} ŋ]	NAS _{SUB}	UNIF-maŋ- _{OTHER}	UNIF-paŋ-RED-	UNIF-maŋ- _{ADV}	UNIF-maŋ-RED-	UNIF-paŋ- _{NOUN}	UNIF-paŋ- _{RES}	<i>weighted sum of constraint violations before noise</i>	<i>weighted sum of constraint violations after noise</i>
[pa-n...] (subst.) 60.5 words	0	1	0	0	0	0	0	0	0	0	1	14.0	13.1 <i>winner</i>
[pan-t...] (unsubst.) 29.5 words	1	0	0	0	1	0	0	0	0	0	0	16.8	16.3

<Tableau for example 14>

<i>constraint</i>	<i>weight</i>
NAS _{SUB}	6.17
*NC̣	10.61
*[_{root} m/n/ŋ]	0.42
*[_{root} n/ ŋ]	4.36
*[_{root} ŋ]	2.17
UNIF-maŋ-OTHER	0.37
UNIF-paŋ-RED	2.12
UNIF-maŋ-ADV	4.31
UNIF-maŋ-RED	5.01
UNIF-paŋ-NOUN	8.94
UNIF-paŋ-RES	13.61

<Tableau for example 15>

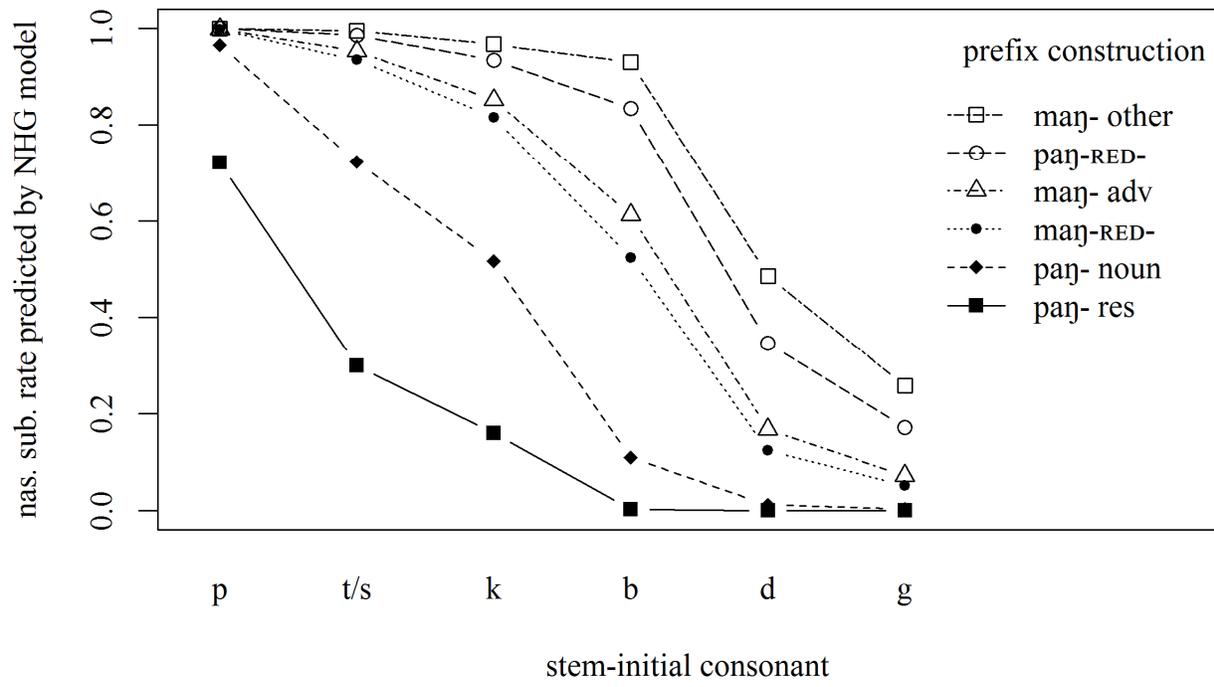


FIGURE 7. Noisy HG grammar's probabilities of nasal substitution

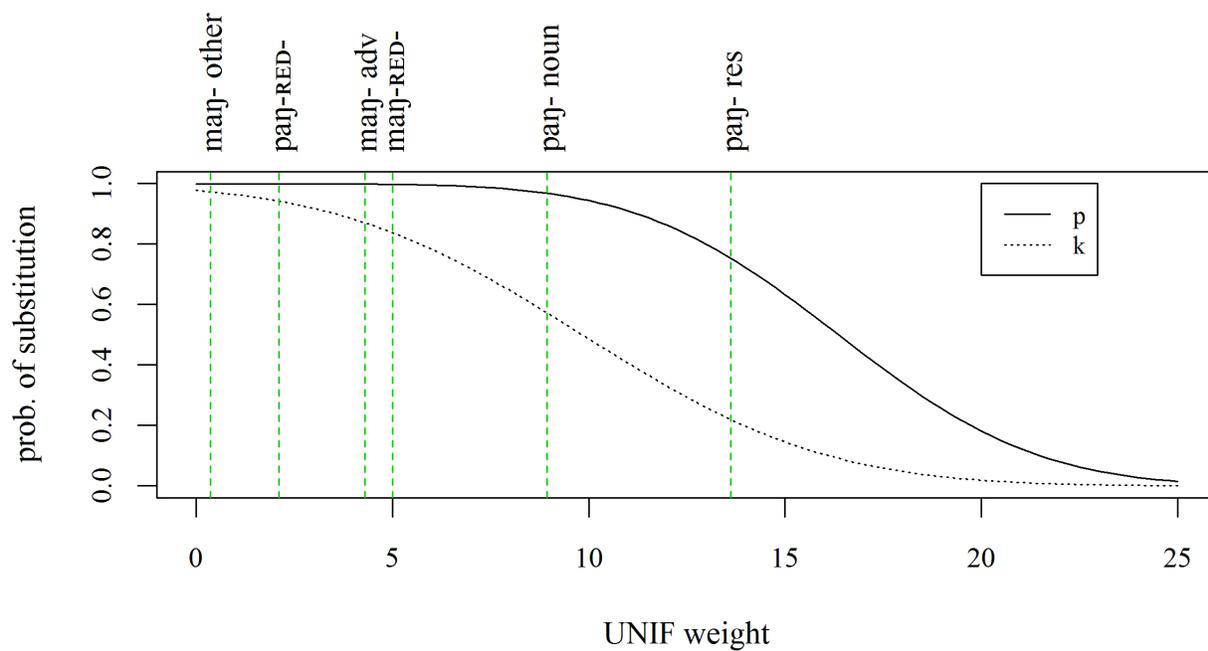


FIGURE 8. Probability of substitution as a function of UNIF weight in Noisy HG

<i>constraint</i>	<i>ranking value</i>
NASUB	-310.72
*NC ₉	-304.15
*[_{root} m/n/ŋ]	-2355.49
*[_{root} n/ ŋ]	-315.55
*[_{root} ŋ]	-308.22
UNIF-man _{-OTHER}	-314.59
UNIF-paŋ-RED	-312.71
UNIF-man _{-ADV}	-309.02
UNIF-man-RED	-309.96
UNIF-paŋ-NOUN	-306.99
UNIF-paŋ-RES	-302.22

TABLE 2. Best GLA-learned Stochastic OT ranking values for Tagalog

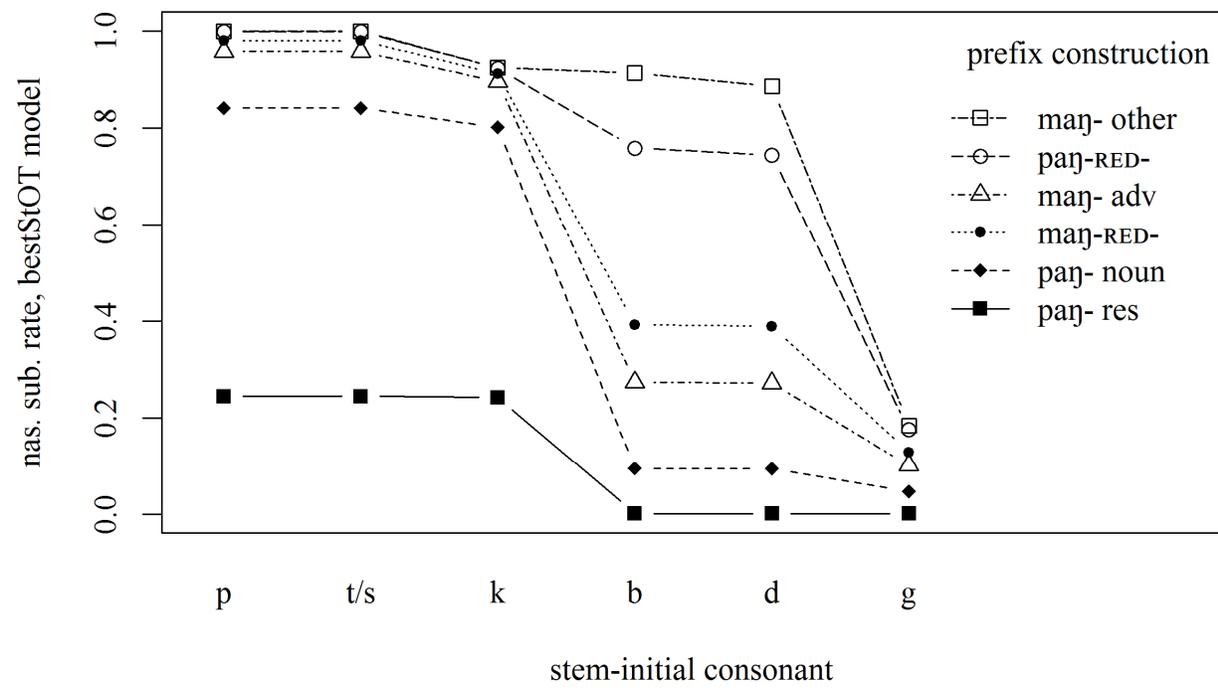


FIGURE 9. Stochastic OT model of Tagalog

<i>constraint</i>	<i>ranking value</i>
*N+p	105.4
*N+t/s	104.2
*N+k	103.1
*N+b	100.1
*N+d	96.1
*N+g	91.0

<i>constraint</i>	<i>ranking value</i>
UNIF-paŋ-RES	105.3
UNIF-paŋ-NOUN	102.5
UNIF-maŋ-RED-	99.9
UNIF-maŋ-ADV	99.2
UNIF-paŋ-RED-	97.1
UNIF-maŋ-OTHER	95.9

TABLE 3. Stochastic OT grammar with consonant-specific constraints: ranking values learned

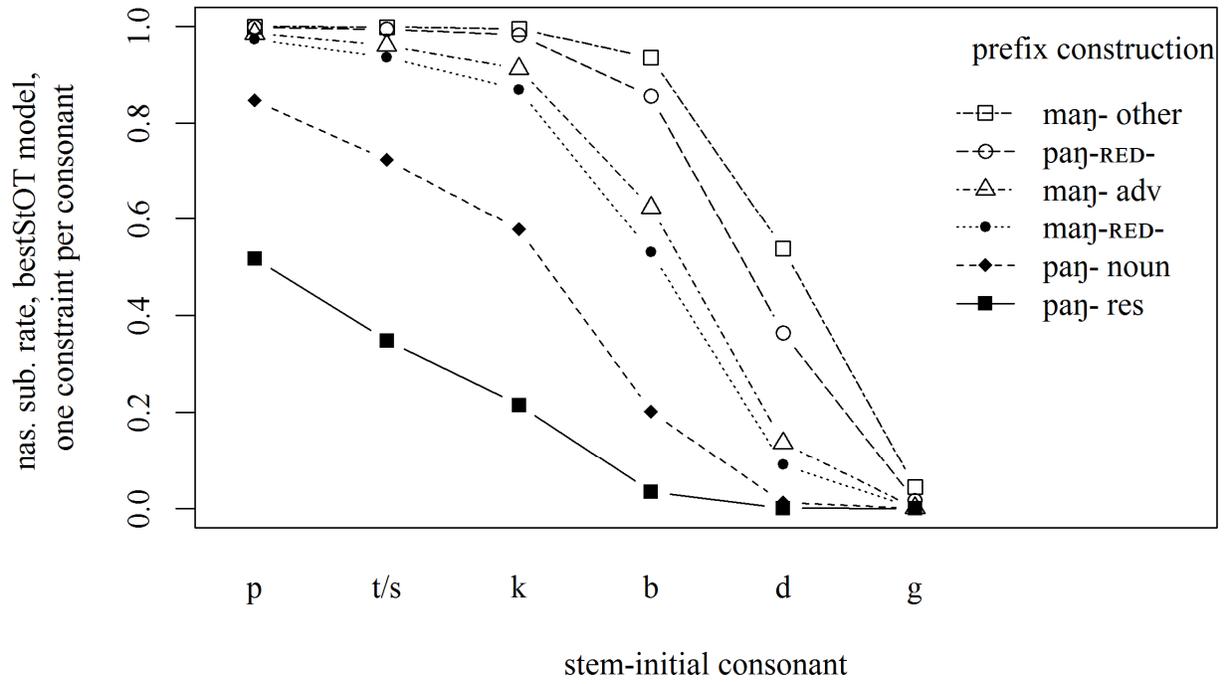


FIGURE 10. Stochastic OT grammar with consonant-specific constraints: predictions

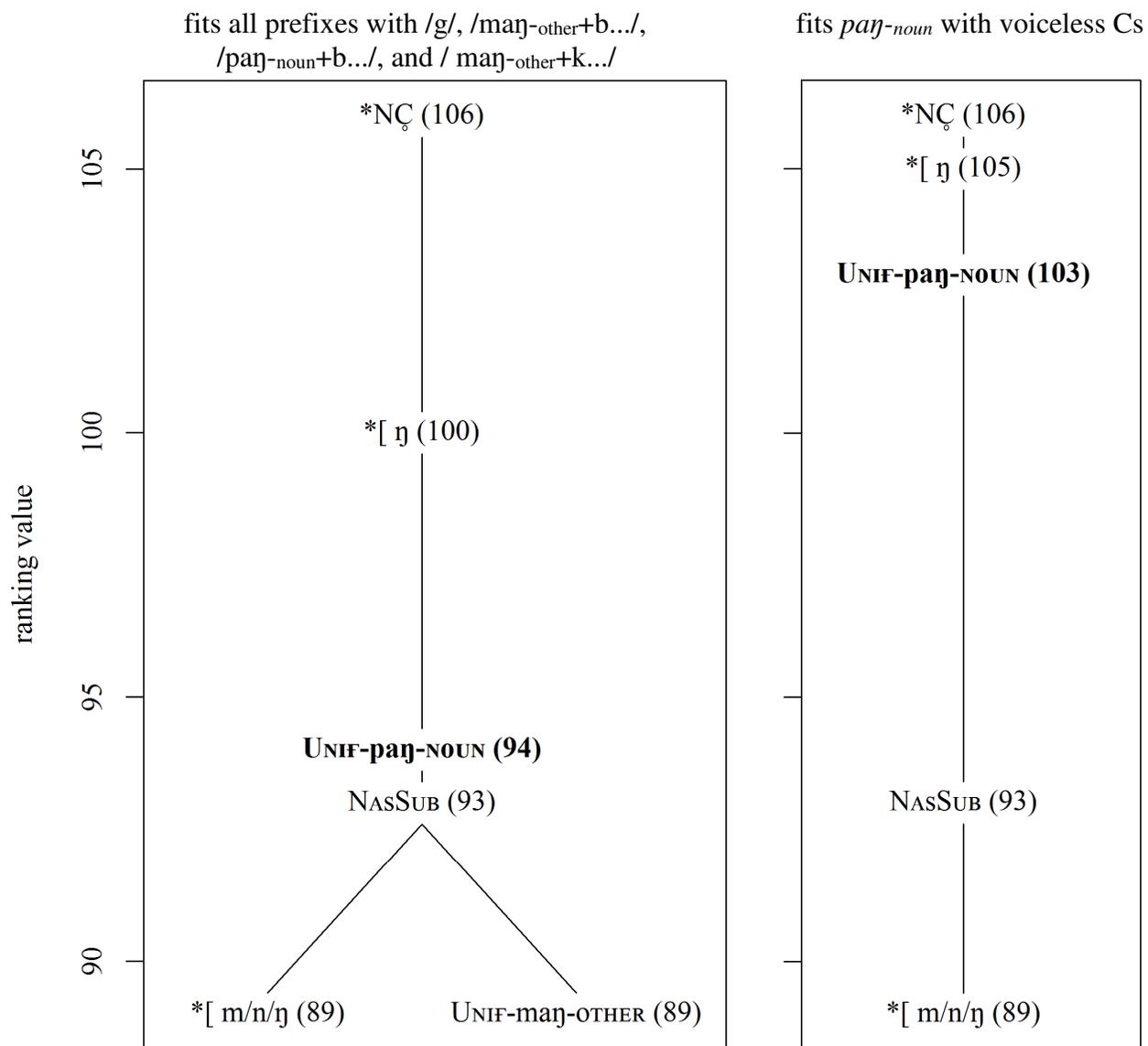


FIGURE 11. Contradictory hand-built Stochastic OT grammar fragments

	<i>constraint</i>	<i>ranking value</i>
Stratum 1	*NÇ	2480
	UNIF-paŋ-NOUN	2480
	UNIF-paŋ-RES	2480
Stratum 2	*[_{root} ŋ]	2440
Stratum 3	*[_{root} n/ŋ]	2400
	NASSUB	2400
	UNIF-maŋ-RED-	2400
Stratum 4	UNIF-maŋ-ADV	2380
Stratum 5	UNIF-paŋ-RED-	1960
Stratum 6	*[_{root} m/n/ŋ]	-2200
Stratum 7	UNIF-maŋ-OTHER	-13400

TABLE 4. Best Stratified Anttilian model

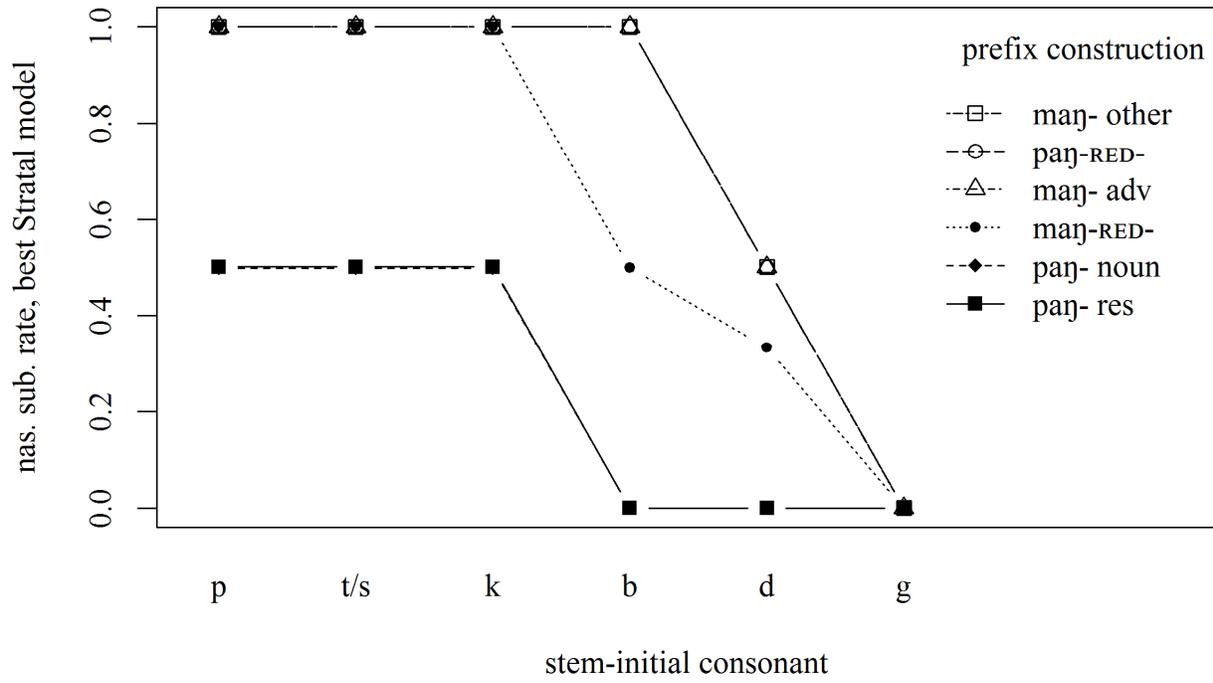
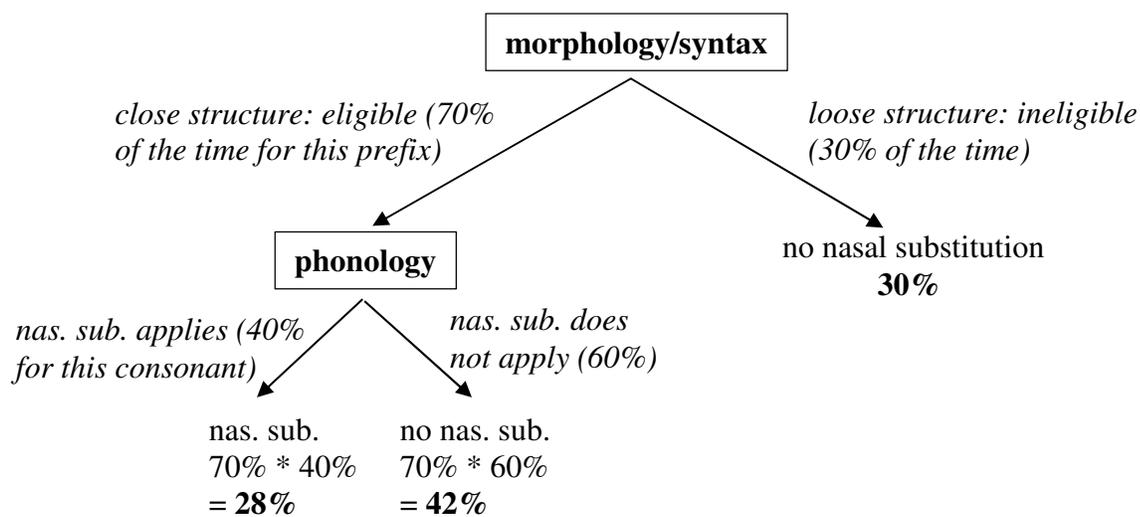


FIGURE 12. Best stratified Anttilian model predictions



Total rate of substitution: 28%

Total rate of non-substitution: 42% + 30% = 72%

FIGURE 13. Decision tree: Morphosyntactic decision feeds phonological decision

<i>morphology</i>	<i>probability of close structure</i>
maŋ-other	1.00
paŋ-RED-	1.00
maŋ-adv	1.00
maŋ-RED-	0.95
paŋ-noun	0.65
paŋ-res	0.27

<i>consonant</i>	<i>probability of substitution, given close structure</i>
p	1.00
t/s	1.00
k	0.98
b	0.74
d	0.21
g	0.01

TABLE 5. Component probabilities fitted for decision tree model

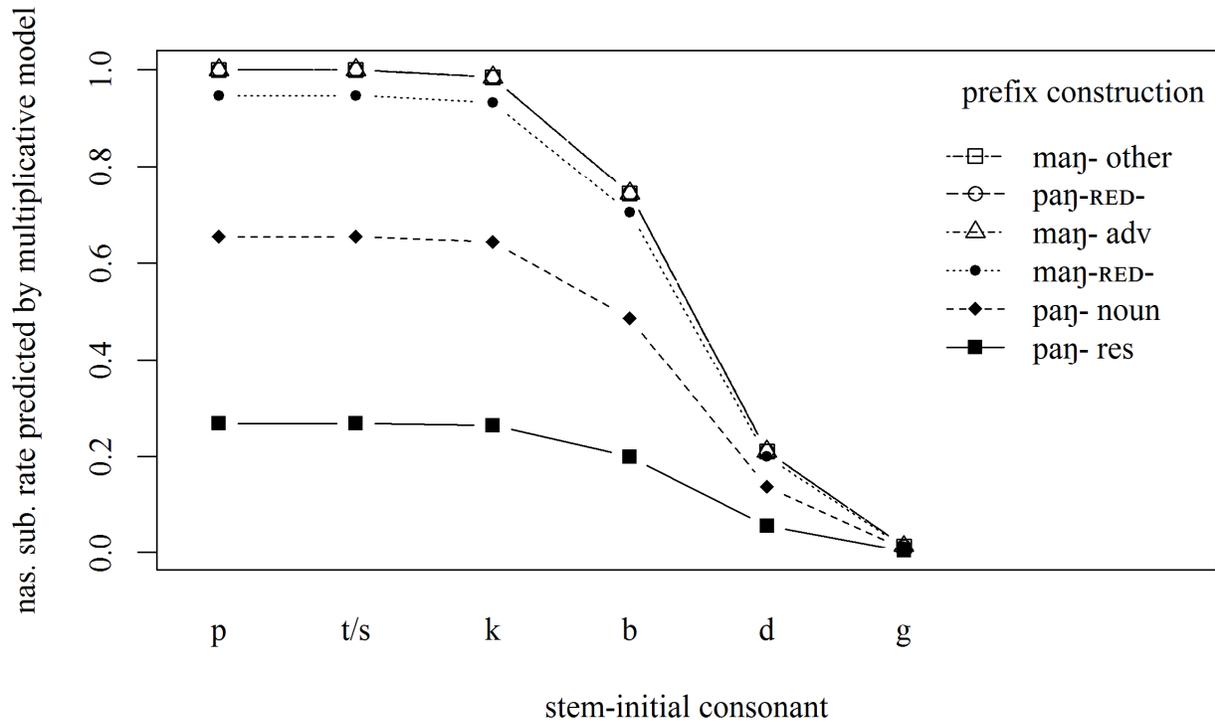


FIGURE 14. Predictions of the decision tree model

	log likelihood closer to 0 is better	number of free parameters
MaxEnt	-284.82 (best)	11
Noisy Harmonic Grammar	-294.48	11
Stochastic OT	-314.64	11
Stratified version of Partial Ordering OT	-645.72	11
Decision tree	-292.31	12
<i>perfect model (correct rate for each C-by-prefix category)</i>	-254.66	36

TABLE 6. Model fits

Word1	default allomorph		liaison/elision allomorph	
‘the- <i>masc</i> ’	<i>le sapin</i> [lə sapɛ̃]	‘the fir tree’	<i>l’if</i> [l if]	‘the yew tree’
‘the- <i>fem</i> ’	<i>la courgette</i> [la kuʁʒɛt]	‘the zucchini’	<i>l’aubergine</i> [l obɛʁʒin]	‘the eggplant’
‘of’	<i>de jonquilles</i> [də ʒɔ̃kiʝ]	‘of daffodils’	<i>d’iris</i> [d iʁis]	‘of irises’
‘of the- <i>masc</i> ’	<i>du petit</i> [dy pətɛ̃]	‘of the small one’	<i>de l’enfant</i> [də l ɑ̃fɑ̃]	‘of the child’
‘at/to the- <i>masc</i> ’	<i>au lac</i> [o lak]	‘at the lake’	<i>à l’étang</i> [a l etɑ̃]	‘at the pond’

TABLE 7. Two allomorphs of Word1

<i>__C form</i>		<i>__V form</i>		
<i>le</i>	[lə]	<i>l'</i>	[l]	'the'- <i>masc</i>
<i>du</i>	[dy]	<i>de l'</i>	[də l]	'of the'- <i>masc</i>
<i>au</i>	[o]	<i>à l'</i>	[a l]	'to the'- <i>masc</i>
<i>de</i>	[də]	<i>d'</i>	[d]	'of'
<i>que</i>	[kə]	<i>qu'</i>	[k]	'that' (complementizer)
<i>se</i>	[sə]	<i>s'</i>	[s]	'it/him/her/one-self'
<i>te</i>	[tə]	<i>t'</i>	[t]	'you'
<i>me</i>	[mə]	<i>m'</i>	[m]	'me'
<i>je</i>	[ʒə]	<i>j'</i>	[ʒ]	'I'
<i>ne</i>	[nə]	<i>n'</i>	[n]	'not'
<i>la</i>	[la]	<i>l'</i>	[l]	'the'- <i>fem</i>
<i>ce</i>	[sə]	<i>cet</i>	[sɛt]	'this'
<i>ma</i>	[ma]	<i>mon</i>	[mɔ̃n]	'my'- <i>fem</i>
<i>ta</i>	[ta]	<i>ton</i>	[tɔ̃n]	'your'- <i>fem</i>
<i>sa</i>	[sa]	<i>son</i>	[sɔ̃n]	'his/her/its'- <i>fem</i>
<i>beau</i>	[bo]	<i>bel</i>	[bɛl]	'pretty'- <i>masc</i>
<i>nouveau</i>	[nuvo]	<i>nouvel</i>	[nuvɛl]	'new'- <i>masc</i>
<i>vieux</i>	[vjø]	<i>vieil</i>	[vjej]	'old'- <i>masc</i>
<i>fou</i>	[fu]	<i>fol</i>	[fɔl]	'crazy'- <i>masc</i>
<i>mou</i>	[mu]	<i>mol</i>	[mɔl]	'limp'- <i>masc</i>

TABLE 8. Full list of Word1s

<i>Alignant</i>			<i>Non-alignant</i>		
<i>le yodle</i>	[lɔ̃ jɔdl]	‘yodels it’	<i>l’iode</i>	[l jɔd]	‘the iodine’
<i>le yaourt</i>	[lɔ̃ jaurɔ̃t]	‘the yogurt’	<i>l’yeuse</i>	[l jøz]	‘the oak’
<i>la hiérarchie</i>	[la jɛʁaʁʁi]	‘the hierarchy’	<i>l’hiatus</i>	[l jatys]	‘the hiatus’
<i>la huée</i>	[la ɥe]	‘the booing’	<i>l’huître</i>	[l ɥitʁ]	‘the oyster’
<i>le huitième</i>	[lɔ̃ ɥitjɛm]	‘the eighth’	<i>l’huile</i>	[l ɥil]	‘the oil’
			<i>l’huissier</i>	[l ɥisje]	‘the bailiff’
<i>le ouistiti</i>	[lɔ̃ wistiti]	‘the marmoset’	<i>l’ouest</i>	[l wɛst]	‘the west’

TABLE 9. Glide-initial words exemplifying each behavior (Walker 2001: 105-106)

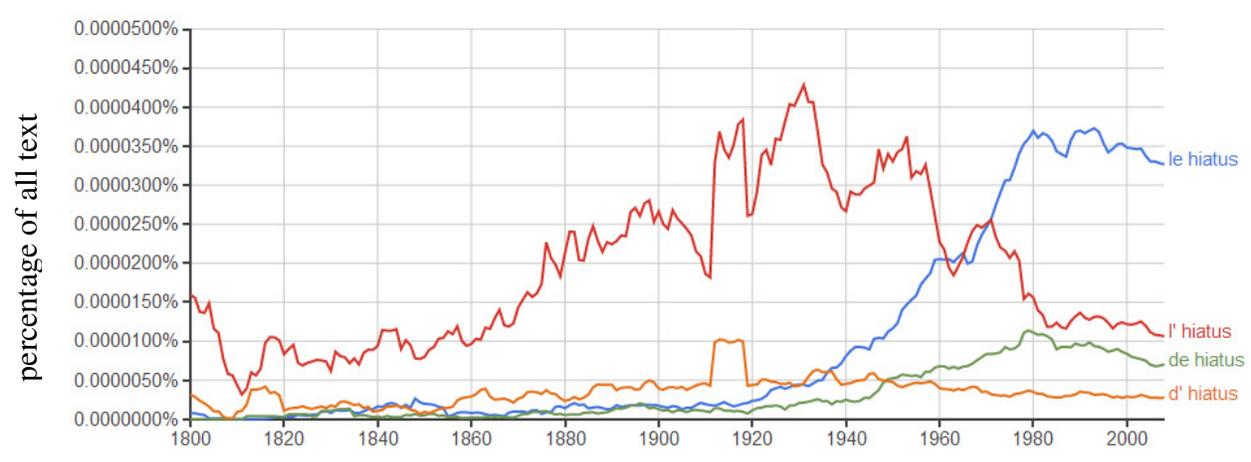


FIGURE 15. Change over time for hiatus 'hiatus'

	/yn(ə) af/	ALIGN	*C.V	DEP-ə/*ə
☞	yn. af		*	
☞	y.n af	*		
☞	y.nə . af			*

<Tableau for example 18>

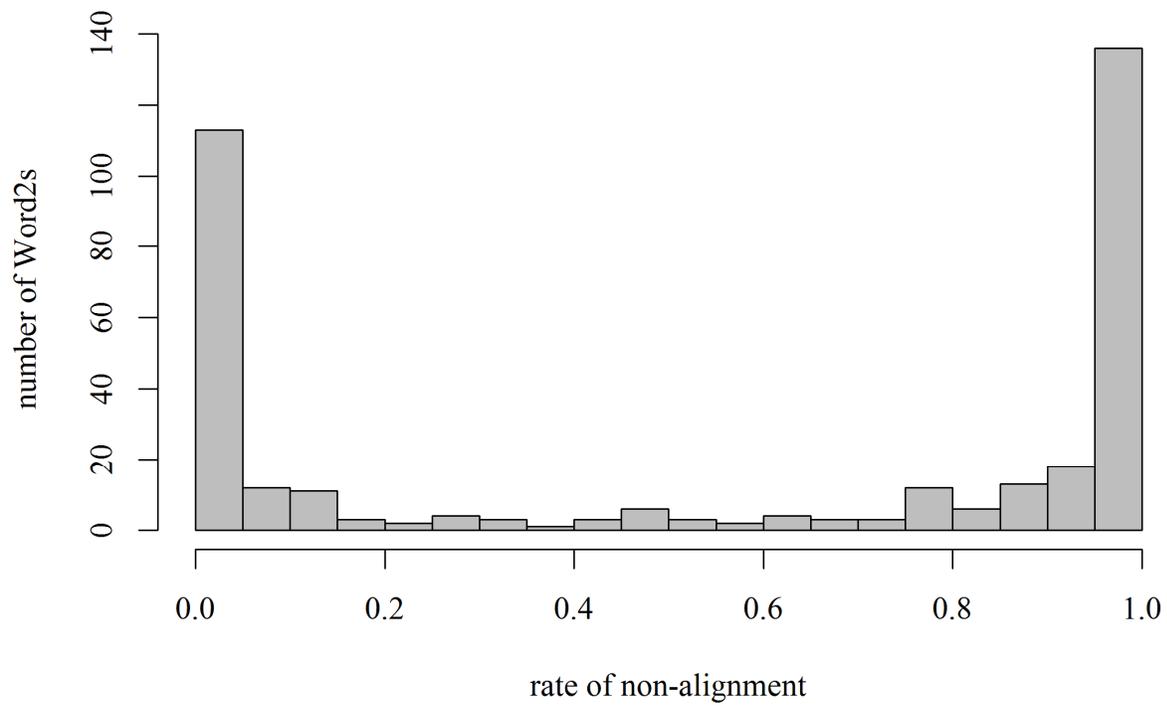


FIGURE 16. Histogram of non-alignant behavior

<i>Word2</i>	<i>random intercept</i>	<i>non-alignment rate</i>	<i>normative behavior</i>	<i>gloss</i>
habituel	2.425	96.9%	non-alignant	'habitual- <i>masc</i> '
habituelle	1.649	99.0%	non-alignant	'habitual- <i>fem</i> '
habitus	2.208	97.1%	non-alignant	'habitus'
hache	-4.861	0.1%	alignant	'axe'
hachette	-3.862	0.0%	alignant	'hatchet; moth sp.'
hacienda	-0.446	78.6%	non-alignant	'hacienda'
haddock	-2.729	0.0%	alignant	'haddock'
Hadès	1.047	85.3%	non-alignant	'Hades'
hadji	-3.052	0.0%	alignant	'haji'
Hadrien	0.633	98.6%	non-alignant	'Hadrian'

TABLE 10. Sample non-alignment index values

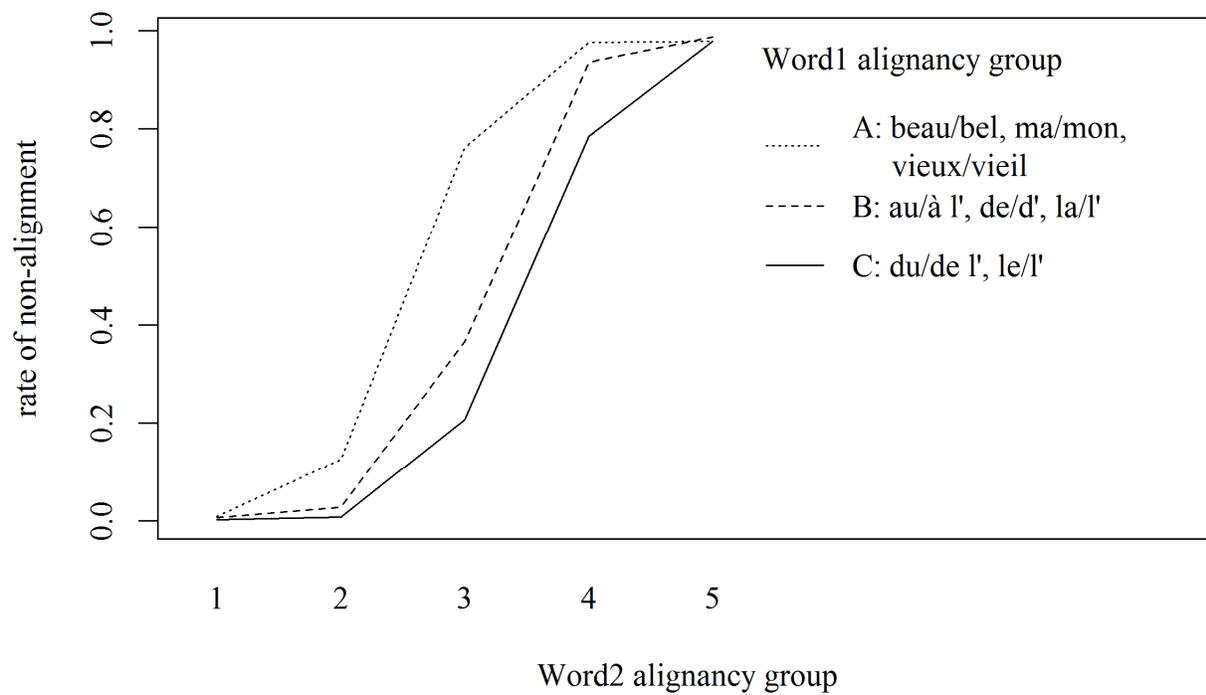


FIGURE 17. Observed non-liaison/élision rate as a function of Word2 and Word1

<i>General phonological constraint</i>	
NOHIATUS	<ul style="list-style-type: none"> • A syllable may not begin with a vowel or glide. • Combines ONSET and the MARGIN family (both Prince & Smolensky 1993/2004). • Promotes non-alignant behavior.
<i>Constraints enforcing propensities of Word1 and Word2</i>	
USEAU USEDU USEDE USELE ...	<ul style="list-style-type: none"> • For each Word1, there is a constraint preferring its citation allomorph. • Promote alignant behavior (use of citation/pre-C form).
ALIGN(MORPHEME, L; SYLL, L)	<ul style="list-style-type: none"> • A morpheme (in particular, Word2), should begin with a syllable boundary (McCarthy & Prince 1993). • There are five versions of this constraint, one for each alignancy group. • Promote alignant behavior (use of citation/pre-C form).

TABLE 11. Constraints for French analysis

/lə i vɛɪ/group 5	ALIGN(Morph, L; Syll, L) group 2	NOHIATUS	USELE	ALIGN(Morph, L; Syll, L) group 5
lə.i.vɛɪ		*!		
☞ li.vɛɪ			*	*

/lə i bu/group 2	ALIGN(Morph, L; Syll, L) group 2	NOHIATUS	USELE	ALIGN(Morph, L; Syll, L) group 5
☞ lə.i.bu		*		
li.bu	*		*	

<Tableau for example 19>

name	MaxEnt weight	Noisy HG weight
ALIGN—group 1	10.13	17.51
ALIGN—group 2	8.01	14.91
ALIGN—group 3	4.95	9.58
ALIGN—group 4	2.06	3.75
ALIGN—group 5	0.00	0.19
NOHIATUS	6.20	12.13
USEMA	0.00	0.01
USEBEAU	0.28	0.22
USEVIEUX	0.88	1.95
USELA	1.34	2.91
USEDE	1.56	3.30
USEAU	2.26	4.61
USELE	2.54	5.07
USEDU	2.78	5.52

TABLE 12. Constraint weights fitted for MaxEnt and Noisy HG models

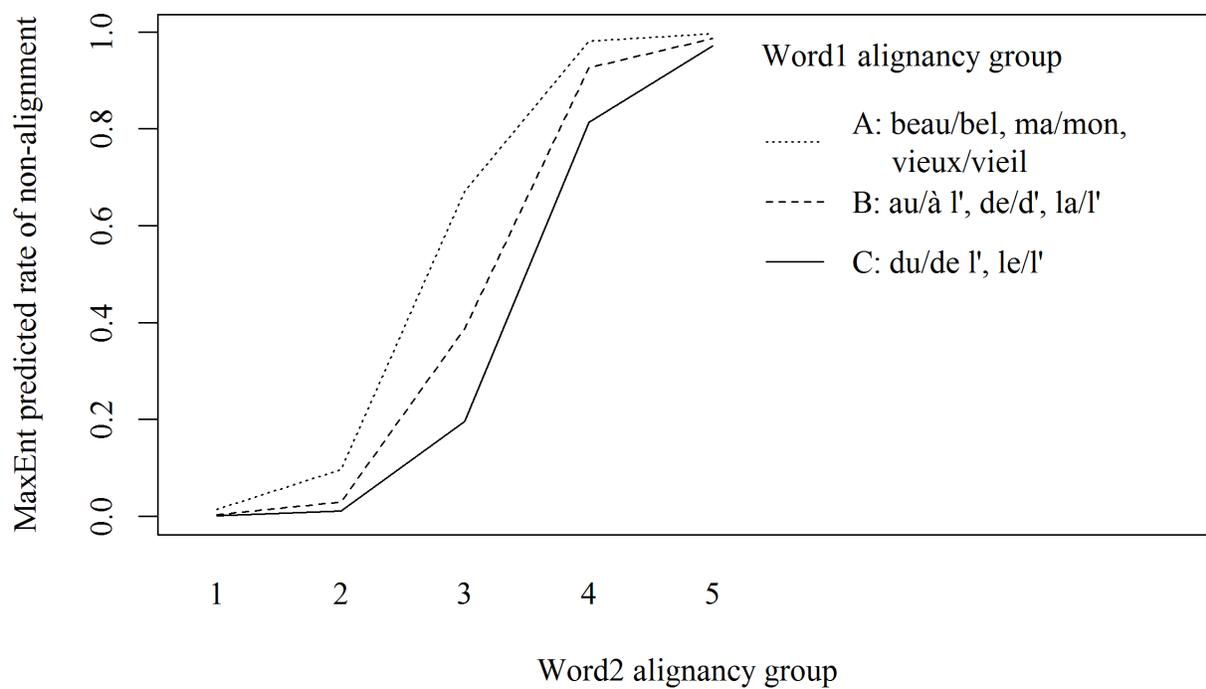


FIGURE 18. MaxEnt model predictions

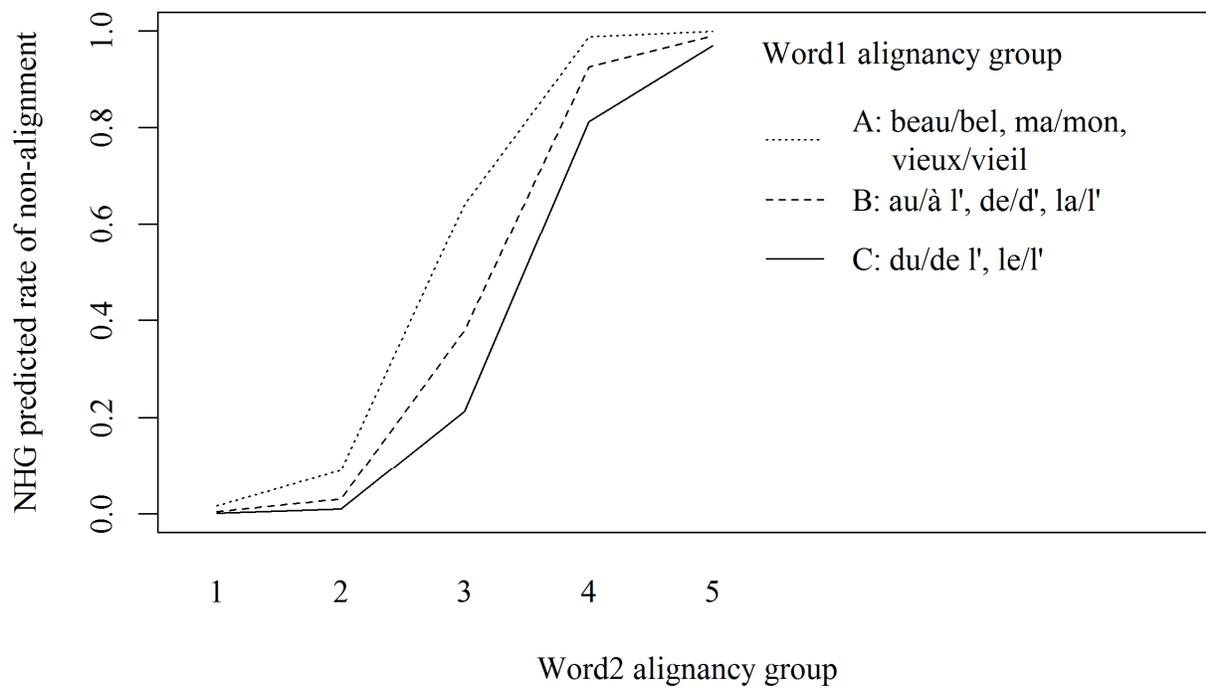


FIGURE 19. Noisy Harmonic Grammar model predictions

<i>constraint</i>	<i>ranking value</i>
ALIGN—group 1	193.52
ALIGN—group 2	192.39
ALIGN—group 3	188.27
ALIGN—group 4	182.86
ALIGN—group 5	-344.23
NOHIATUS	187.18
USEMA	-34.52
USEBEAU	11.77
USEVIEUX	21.09
USELA	76.75
USEDE	90.67
USEAU	181.11
USELE	182.89
USEDU	183.06

TABLE 13. Ranking values of best Stochastic OT model

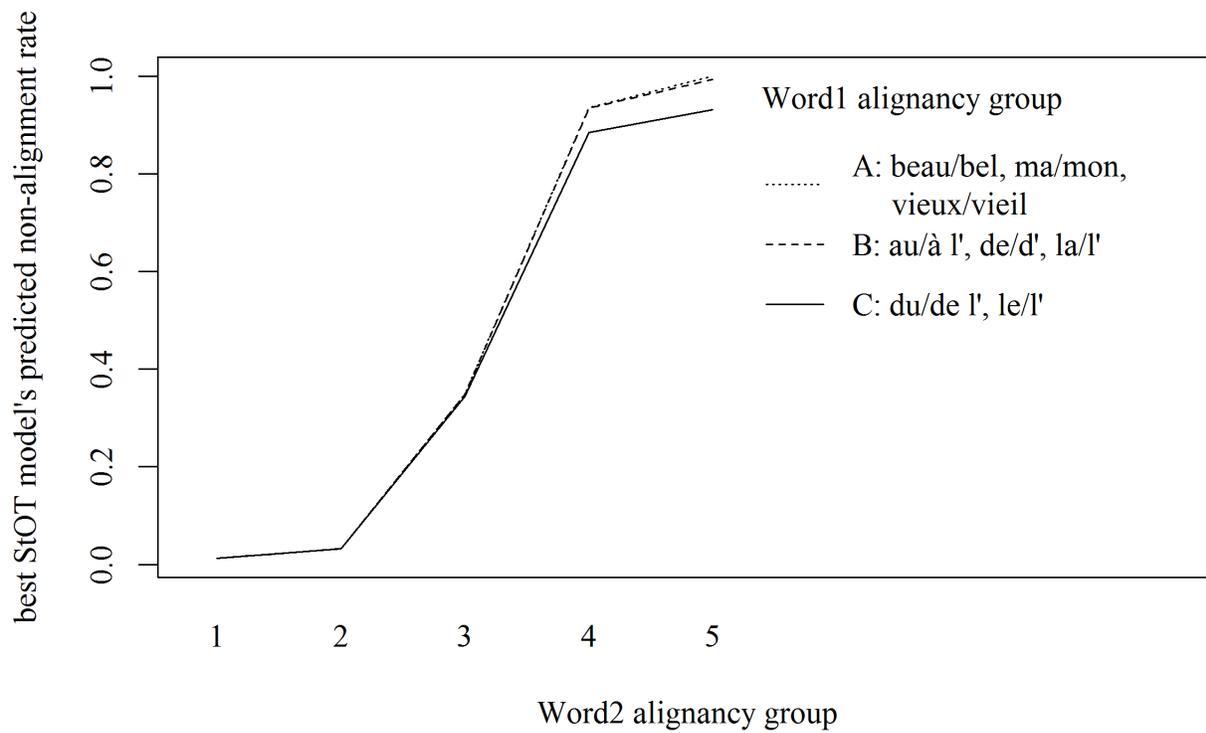


FIGURE 20. Predictions of best Stochastic OT model

	<i>constraint</i>	<i>ranking value</i>
Stratum 1	ALIGN—group 1	3460
	ALIGN—group 2	3460
Stratum 2	ALIGN—group 3	3420
	ALIGN—group 4	3420
	NOHIATUS	3420
Stratum 3	USEAU	3400
	USELE	3400
Stratum 4	USEDU	3380
Stratum 5	USELA	-1260
Stratum 6	USEDE	-1820
Stratum 7	USEVIEUX	-1880
Stratum 8	USEBEAU	-3100
Stratum 9	USEMA	-4640
Stratum 10	ALIGN—group 5	-16580

TABLE 14. Best Stratified Partial Ordering OT model

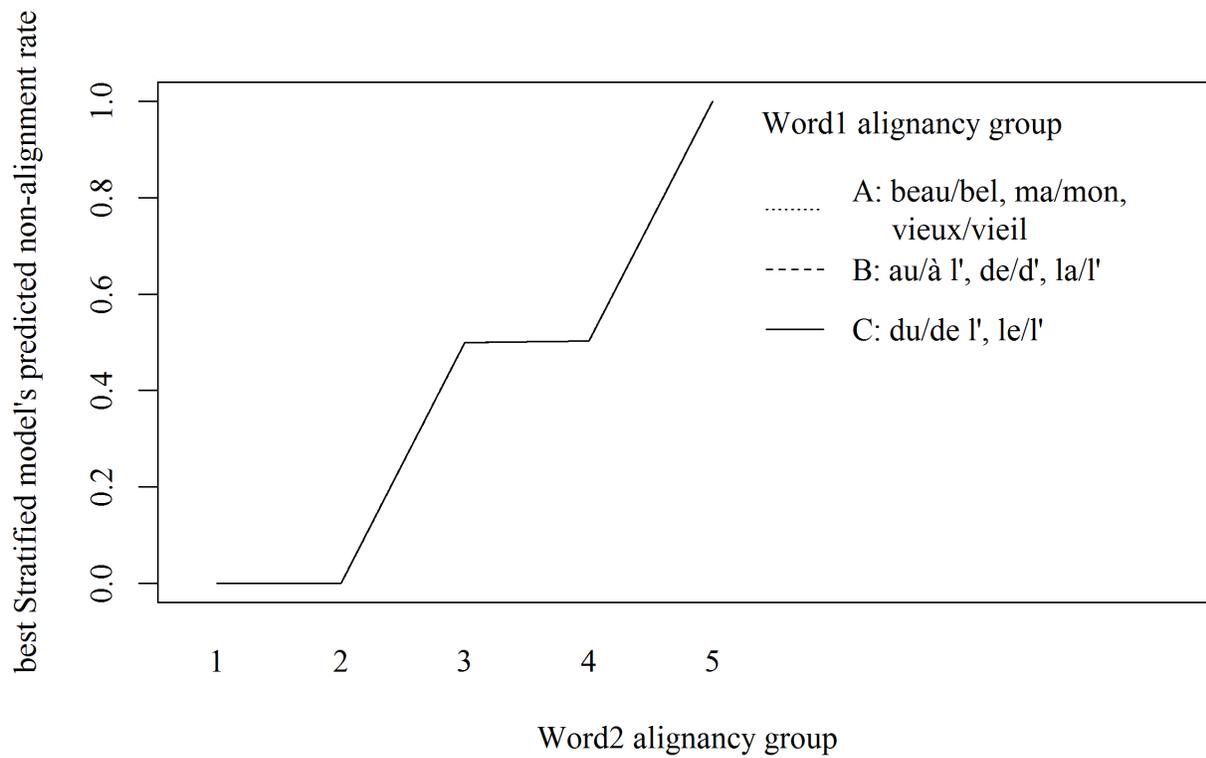


FIGURE 21. Predictions of best stratified Partial Ordering OT model

<i>Word1</i>	<i>probability of close structure</i>
beau/bel	1.00
ma/mon	1.00
vieux/vieil	0.96
au/à l'	1.00
de/d'	0.98
la/l'	1.00
du/de l'	0.97
le/l'	0.96

<i>Word2 group</i>	<i>probability of non-alignment, given close structure</i>
1	0.01
2	0.03
3	0.34
4	0.94
5	1.00

TABLE 15. Component probabilities fitted for decision tree model

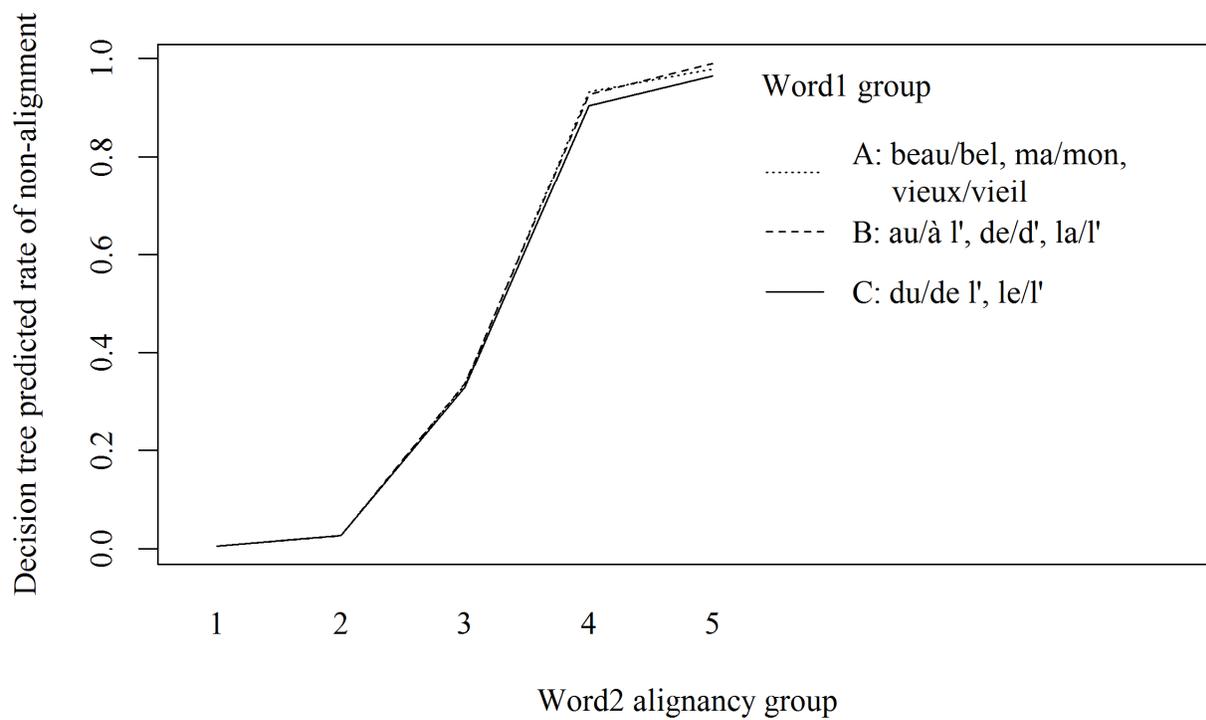


FIGURE 22. Results of fitting decision tree

model	log likelihood (closer to 0 is better)	number of free parameters
MaxEnt	-197.71 (best)	14
Noisy Harmonic Grammar	-198.80	14
Stochastic OT (best fit)	-233.61	14
Stratified Partial Ordering OT (best fit)	-410.64	14
Decision tree	-207.95	13
<i>perfect model (baseline)</i>	<i>-189.36</i>	<i>40</i>

TABLE 16. Comparison of model fits for French

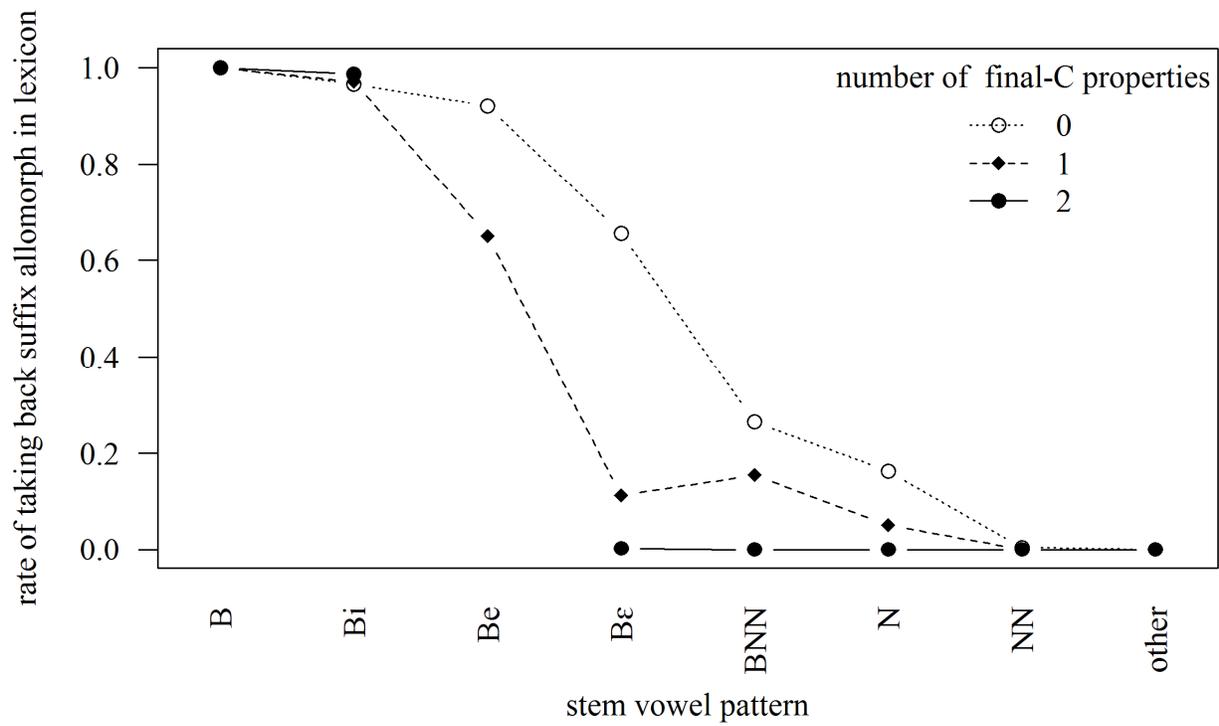


FIGURE 23. Hungarian lexical data

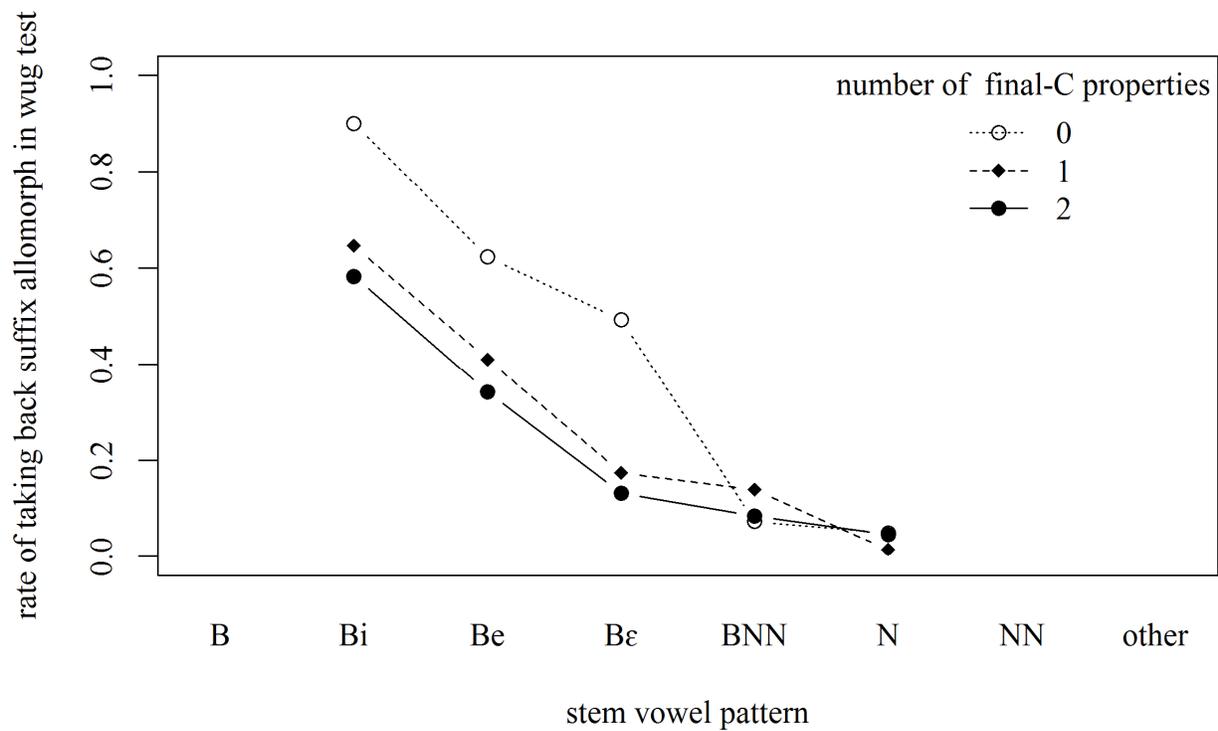


FIGURE 24. Hungarian nonce-stem data

I. Vowel family

<i>Constraint</i>	<i>schematically, in features</i>	<i>Comment</i>
a. AGREE(back, local)	*[+back][−back]	Vowels must agree in backness with a preceding back vowel.
b. AGREE(back, nonlocal)	*[+back] N ₀ [−back]	Nonlocal version of (a)
c. AGREE(front rounded)	*[−back,+round] N ₀ [+back]	Vowels must agree in backness with a preceding front rounded vowel, even if neutral vowels intervene.
d. AGREE(front, local)	*[−back][+back]	Vowels must agree in backness with a preceding front vowel.
e. AGREE(non-high front, local)	*[−back,−high][+back]	Specific version of (e), limited to non-high triggers
f. AGREE(low front, local)	*[−back,+low][+back]	Specific version of (f), limited to low triggers ¹
g. AGREE(double front, local) ²	*[−back][−back][+back]	Two-trigger harmony
h. MONOSYLL-[i:]: prefer back suffixes when the stem is monosyllabic with the vowel [i:].		An unnatural constraint; relevant stems arise from historical back unrounded vowels.

II. Consonant family

- a. BILABIAL: Prefer front suffixes when the stem ends in a bilabial noncontinuant ([p, b, m]).
- b. SIBILANT: Prefer front suffixes when the stem ends in a sibilant ([s, z, ʃ, ʒ, ts, tʃ, dʒ]).
- c. CORSON: Prefer front suffixes when the stem ends in a coronal sonorant ([n, ɲ, l, r]).
- d. CLUSTER: Prefer front suffixes when the stem ends in a sequence of two consonants.

TABLE 17. Constraints for Hungarian

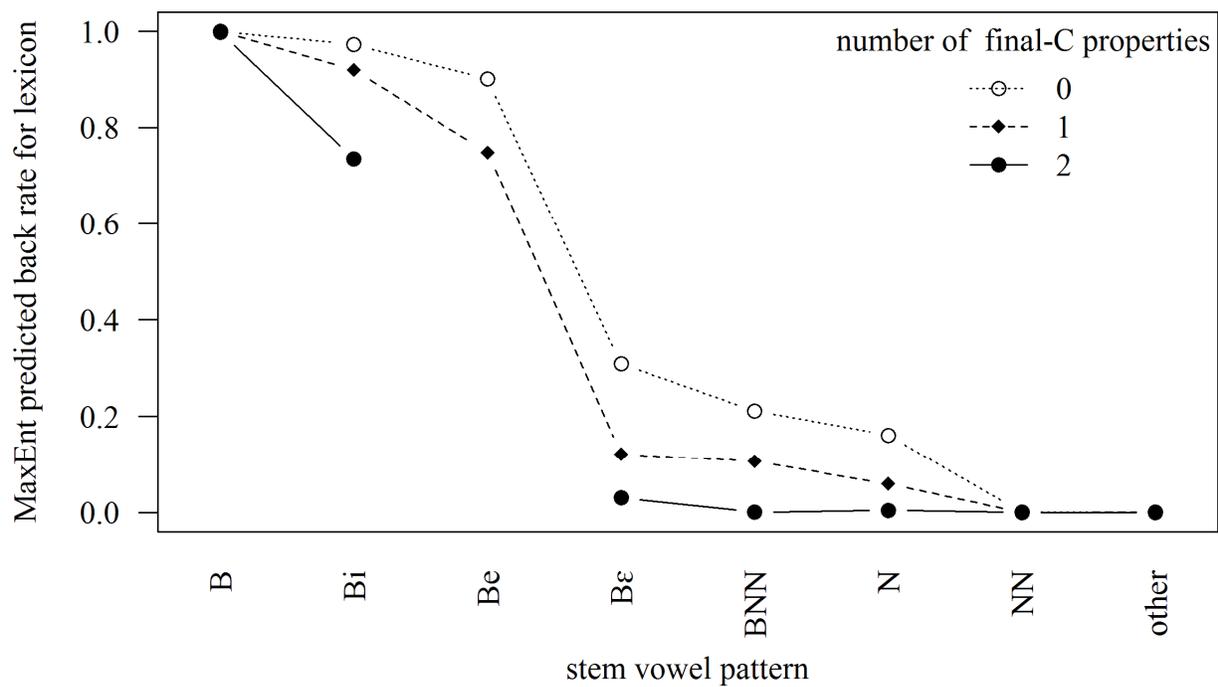
¹ We treat [ɛ], the lowest front vowel of Hungarian, as [+low]: in suffixes it alternates with a low vowel, and phonetically it is often rather lower than the IPA symbol suggests.

² We assign two violations to BNNN-*nɔk* and NNN-*nɔk*, and three violations to NNNN-*nɔk*.

<i>Constraint</i>	<i>MaxEnt weight</i>	<i>Noisy HG weight</i>
AGREE(back, local)	3.96	6.18
AGREE(back, nonlocal)	5.23	7.41
AGREE(front rounded)	4.03	3.32
AGREE(front, local)	1.66	1.59
AGREE(non-high front, local)	1.37	1.71
AGREE(low front, local)	3.01	6.72
AGREE(double front, local)	3.77	6.53
MONOSYLL-[i:]	2.36	2.27
BILABIAL	2.38	3.34
SIBILANT	0.84	0.65
CORSON	1.04	1.02
CLUSTER	1.71	2.28

TABLE 18. Constraint weights fitted for MaxEnt and Noisy HG models

a. Predictions for lexical data



b. Predictions for nonce-probe data

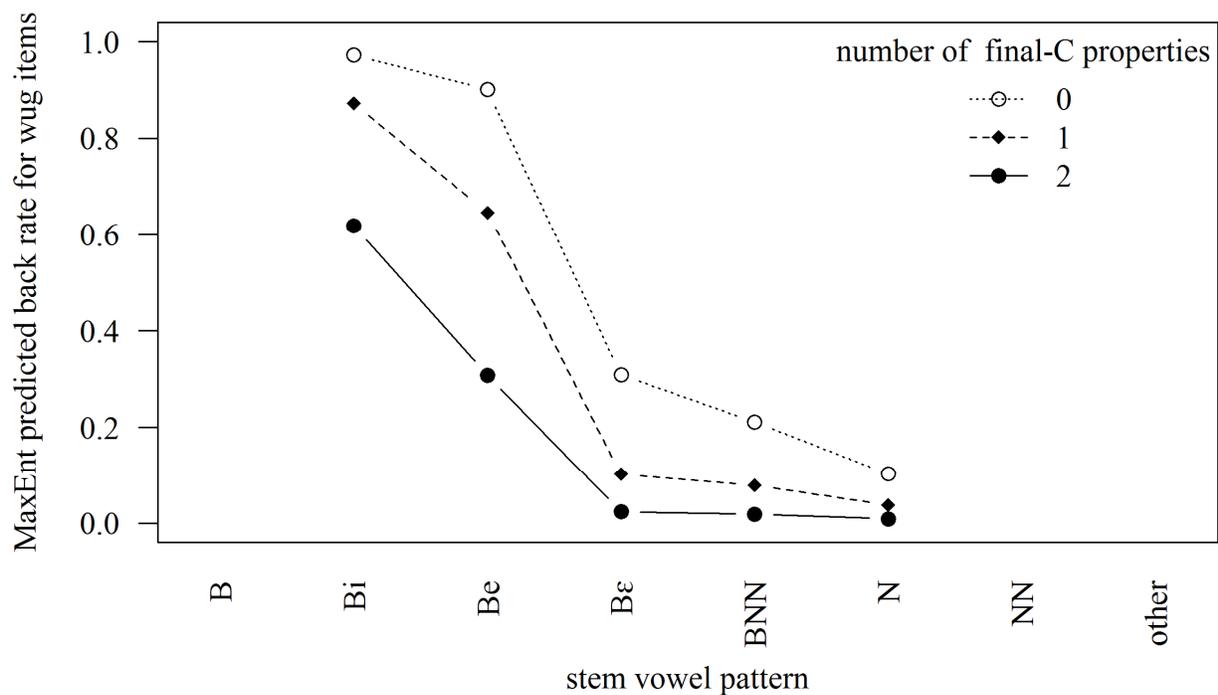


FIGURE 25. Maximum Entropy model predictions for Hungarian

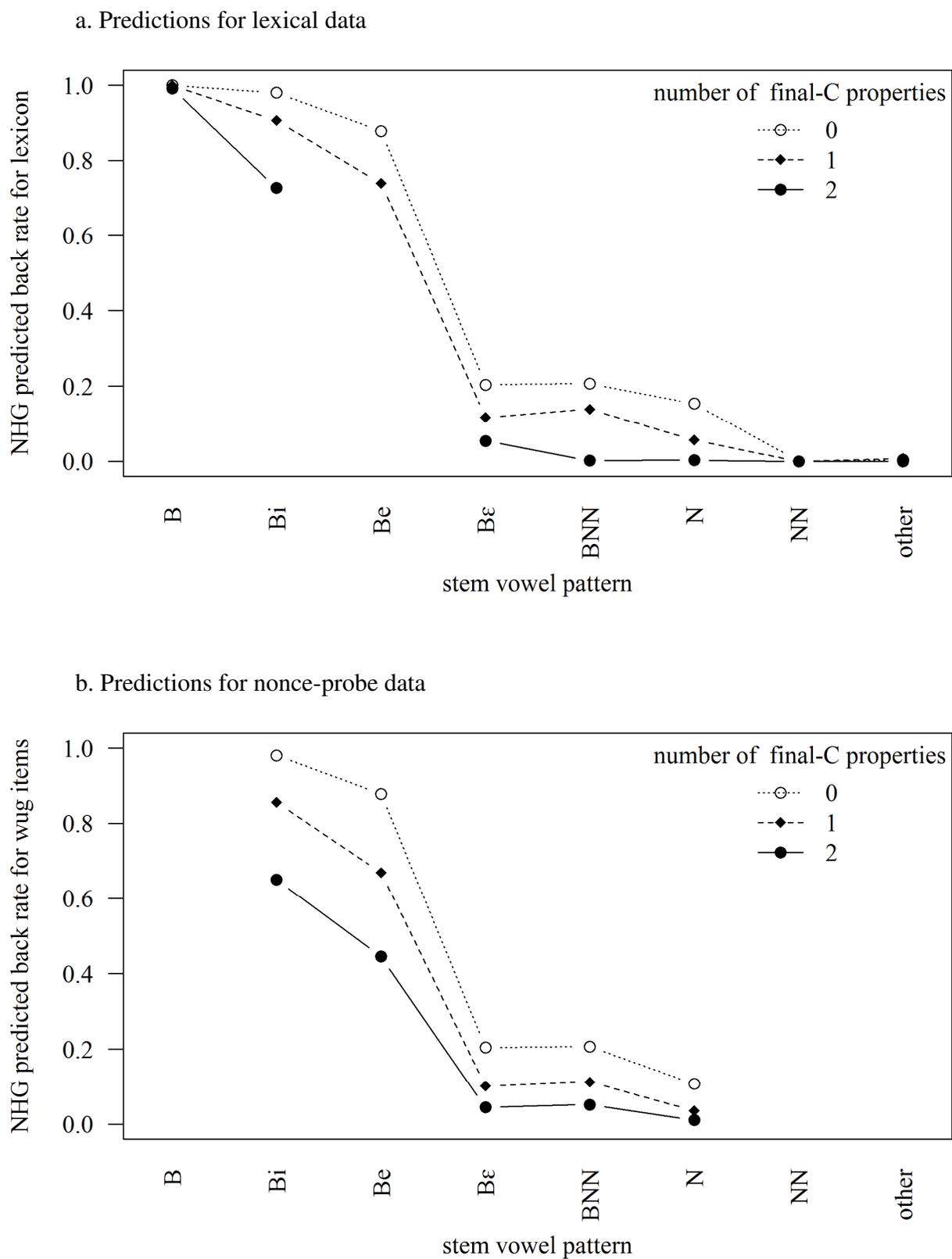


FIGURE 26. Noisy Harmonic Grammar model predictions for Hungarian

<i>Constraint</i>	<i>StOT ranking value</i>
AGREE(back, local)	89.90
AGREE(back, nonlocal)	82.75
AGREE(front rounded)	85.97
AGREE(front, local)	-214.14
AGREE(non-high front, local)	79.58
AGREE(low front, local)	86.38
AGREE(double front, local)	84.78
MONOSYLL-[i:]	76.24
BILABIAL	82.07
SIBILANT	78.21
CORSON	78.24
CLUSTER	80.69

TABLE 19. Constraint ranking values fitted for Stochastic OT model

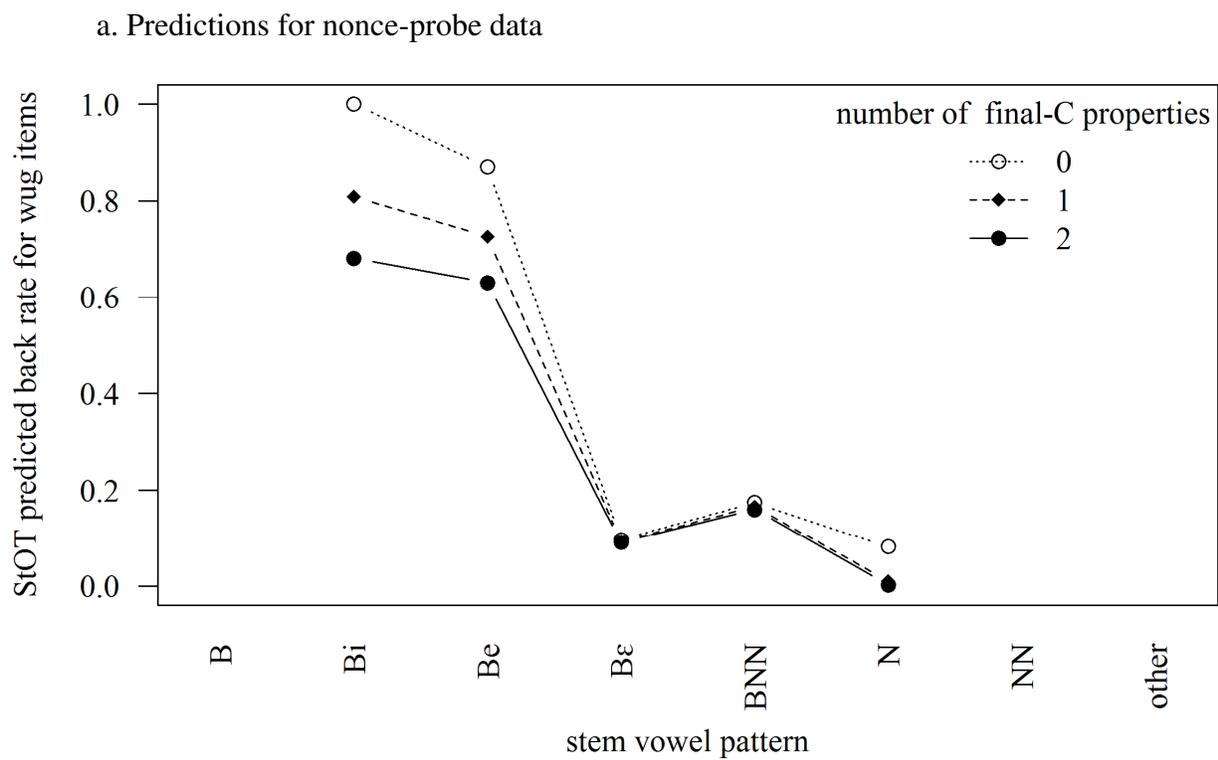
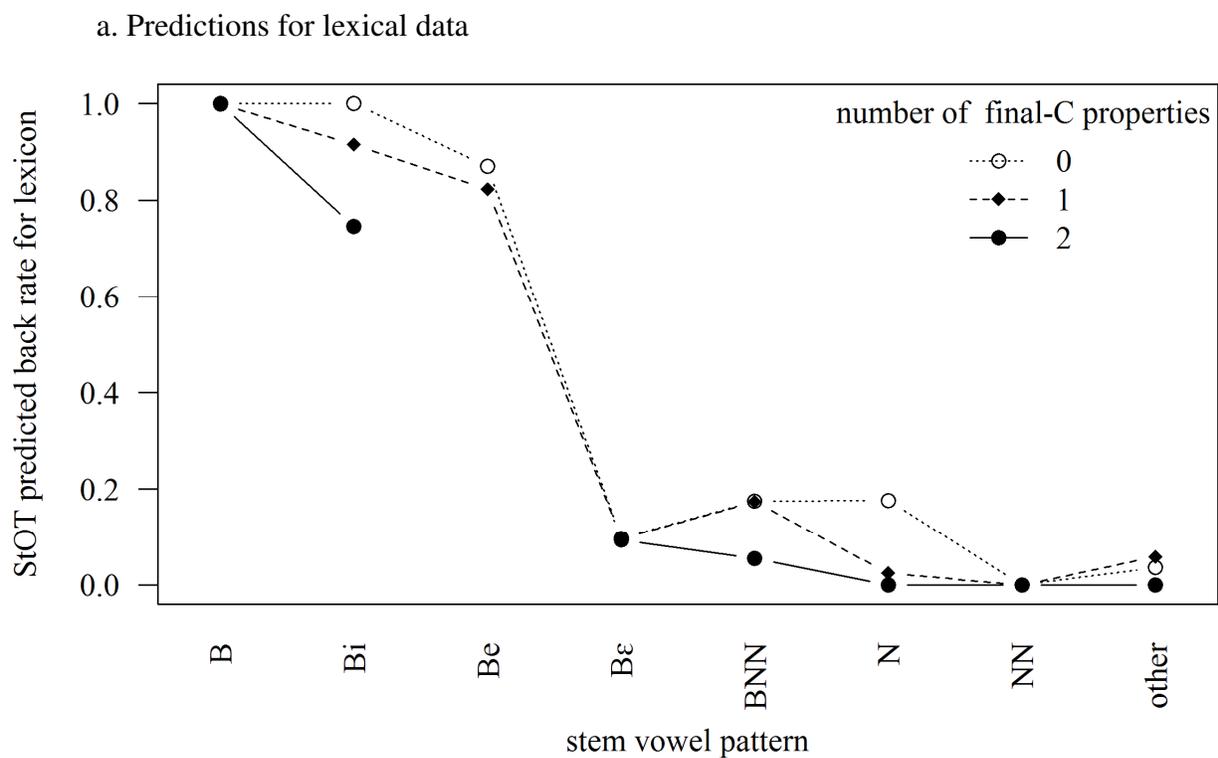
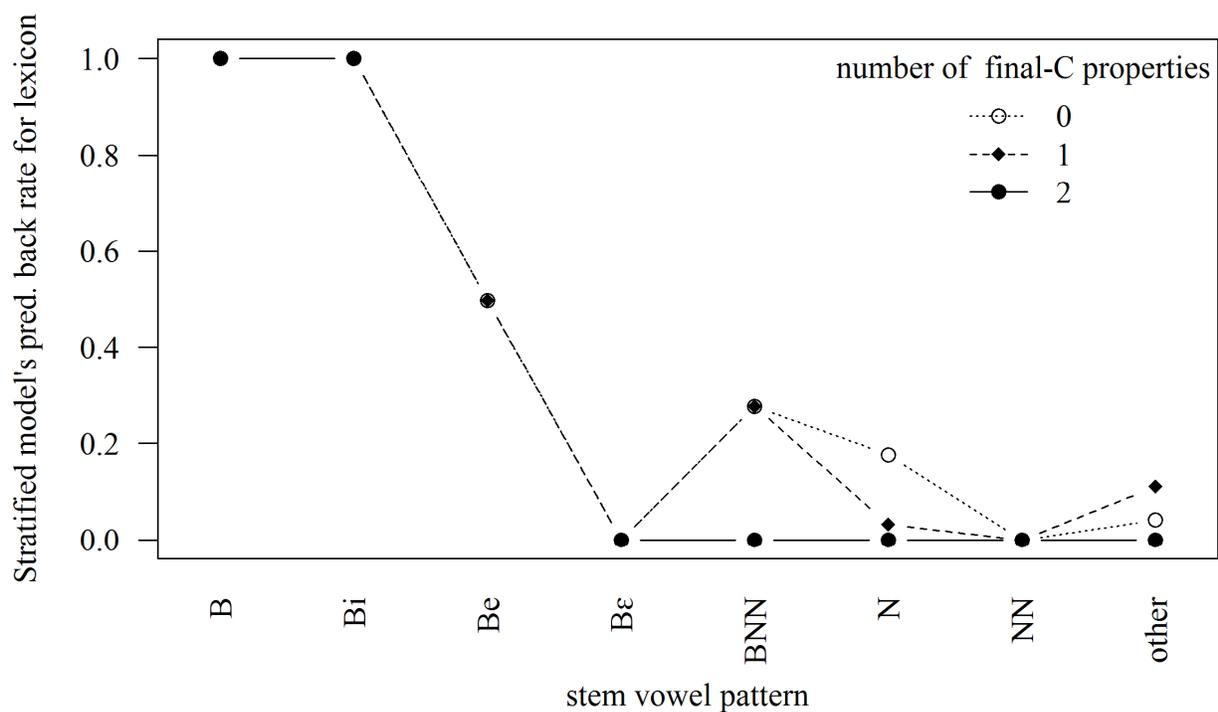


FIGURE 27. Stochastic OT model predictions for Hungarian

	<i>Constraint</i>	<i>ranking value</i>
Stratum 1	AGREE(front rounded)	-100
Stratum 2	AGREE(back, local)	-880
	AGREE(low front, local)	-880
Stratum 3	AGREE(back, nonlocal)	-900
	AGREE(non-high front, local)	-900
	AGREE(double front, local)	-900
Stratum 4	BILABIAL	-920
	CLUSTER	-920
Stratum 5	SIBILANT	-940
Stratum 6	MONOSYLL-[i:]	-980
	CORSON	-980
Stratum 7	AGREE(front, local)	-10620

TABLE 20. Best Stratified Partial Ranking OT mode

a. Predictions for lexical data



a. Predictions for nonce-probe data

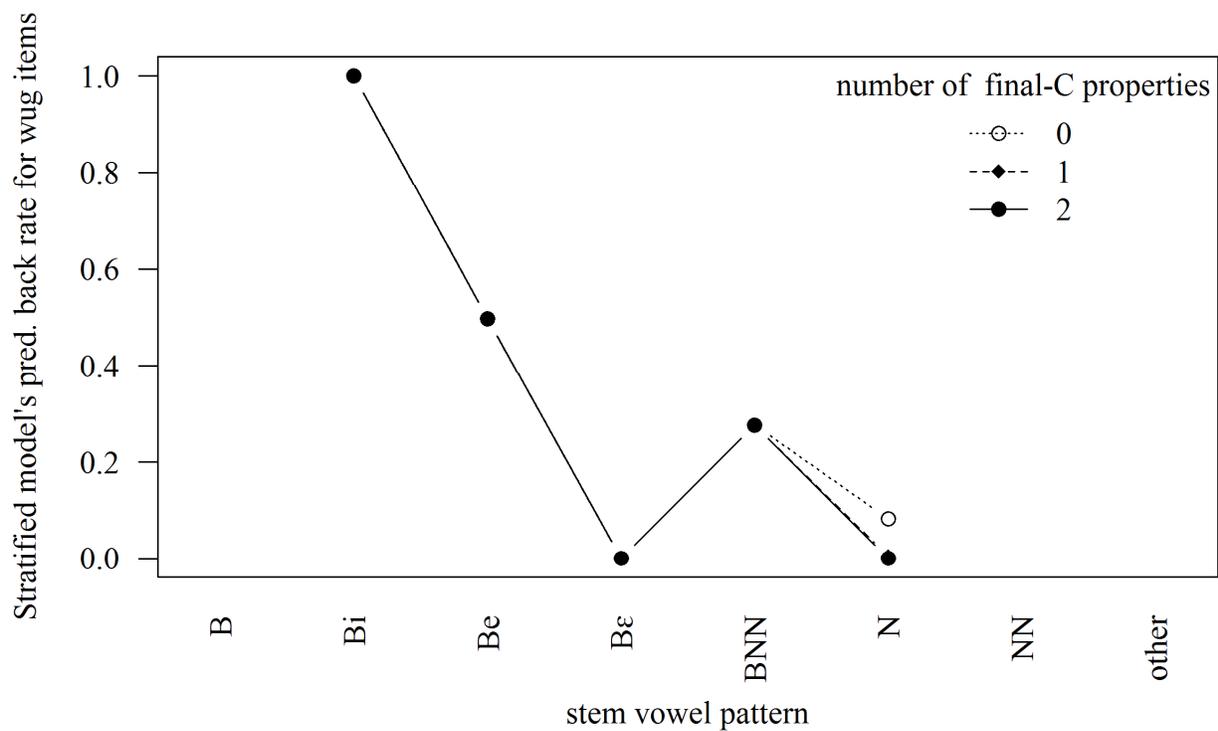


FIGURE 28. Stratified Partial Ordering OT model's predictions for Hungarian

Model	log likelihood for lexical data (closer to 0 is better) <i>n</i> of stems = 8,915	log likelihood for wug data (closer to 0 is better) <i>n</i> of stems = 1,602	number of free parameters
MaxEnt	-187.06 (best)	-823.28 (tied for best)	12
Noisy Harmonic Grammar	-196.76	-823.29 (tied for best)	12
Stochastic OT (best fit)	-280.83	-971.28	12
Stratified Partial Ordering OT (best fit)	-432.68	-2685.75	12
<i>perfect model (baseline)</i>	<i>-164.06</i>	<i>-648.50</i>	<i>232</i>

TABLE 21. Comparison of model fits for Hungarian