Phonology **33** (2016) 125–167. © Cambridge University Press 2016.
doi:10.1017/S0952675716000051

Relating application frequency to morphological structure: the case of Tommo So vowel harmony[*]

Laura McPherson
Dartmouth College

Bruce Hayes
University of California, Los Angeles

We describe three vowel-harmony processes in Tommo So and their interaction with morphological structure. The verbal suffixes of Tommo So occur in a strict linear order, establishing a Kiparskian hierarchy of distance from the root. This distance is respected by all three harmony processes; they 'peter out', applying with lower frequency as distance from the root increases. The function relating application rate to distance is well fitted by families of sigmoid curves, declining in frequency from one to zero. We show that, assuming appropriate constraints, such functions are a direct consequence of Harmonic Grammar. The crucially conflicting constraints are IDENT (violated just once by harmonised candidates) and a scalar version of AGREE (violated one to seven times, based on closeness of the target to the root). We show that our model achieves a close fit to the data, while a variety of alternative models fail to do so.

1 Introduction

In Tommo So, a Dogon language of Mali, there are three vowel-harmony processes, all optional. Our focus is on the fact that, for all three processes,

---

[*] E-mail: LAURA.E.MCPHERSON@DARTMOUTH.EDU, BHAYES@HUMNET.UCLA.EDU.

We would like to thank Robert Daland, Abbie Hantgan, Jeffrey Heath, Kie Zuraw, audiences at the Linguistic Society of America, the Manchester Phonology Meeting, Academia Sinica, Dartmouth College and UCLA, and the anonymous reviewers and editors for *Phonology* for their help with this article; all remaining defects are the authors' responsibility. We are also indebted to the Tommo So language consultants, who provided all of the data. We would also like to thank Joni Ricks and her colleagues at the Statistical Consulting Group of the UCLA Institute for Digital Research and Education for their guidance in preparing and interpreting the statistical tests reported here. The materials used for all the tests (including the program code for the Monte Carlo simulation) are available as online supplementary materials at http://www.journals.cambridge.org/issue_Phonology/Vol33No01. This research was supported by a grant from the Fulbright Foundation and by grants BCS-0537435 and BCS-0853364 from the U.S. National Science Foundation.

frequency of application interacts with morphological structure: intuitively, harmony 'peters out', in the sense that it applies with gradually diminishing frequency in outer layers of the morphology (defined by the possibilities of affix order, following Kiparsky 1982). Each process peters out at a different rate: Low Harmony first (applying only within the innermost morphological layers), Backness Harmony next and ATR Harmony last. The graph in Fig. 1 gives the basic pattern.

We propose a formal model of this pattern, drawing on two ideas in phonological theory: Harmonic Grammar and scalar constraints. We show that this model achieves a far better fit to the data than a number of alternatives, notably the inverted exponential model of Guy (1991).

Our paper is structured as follows: in §2 we review a crucial generalisation from Kiparsky that establishes the theoretical context of our work. §3 gives basic background (the vowel inventory and verbal morphology of Tommo So), then motivates the morphological layers with data from affix ordering. In §4, we illustrate the three vowel-harmony processes, and give the quantitative data. §5 applies Guy's multiplicative model to our data and demonstrates that it provides a poor fit. §6 is the main analytic section: we propose a set of constraints and deploy them in Harmonic Grammar, achieving a greatly improved fit to the data. We show that our approach is restrictive: only certain frequency patterns can be generated. We also review other linguistic phenomena showing similar quantitative patterns. §7 addresses theories that can describe our data but are not restrictive. §8 covers residual issues, and §9 concludes.

2 The theoretical context: affix order and phonological process applicability

Our work pursues an insight from Kiparsky (1982): affix ordering and the applicability of phonological processes tend to be closely correlated. Specifically, affixes that occur 'closer to the root', as diagnosed by ordering tests, characteristically trigger or undergo more phonological processes. Kiparsky suggests that this correlation is 'a general property of languages' (1982: 11), and in light of this proposes a Strong Domain Hypothesis (1984: 142): 'at lower levels of the lexicon and in the postlexical phonology rules may be 'turned off' but no new ones may be added'. Our Tommo So data follow this pattern, but in a gradient fashion ('turning off' is gradual); thus we will be treating the Kiparskian correlation in different theoretical terms.

We emphasise that the concept of root closeness on which Kiparsky relies is abstract: it is *not* the literal distance seen within individual forms, but is calculated by examining the morphology as a system. To give Kiparsky's example (1982: 11): English has two negative prefixes, *non-* and *in-*, of which the latter may occur as *il-* by assimilation, as in *illegible*. When the two prefixes co-occur, *non-* may precede *in-*, but not *vice versa: non-illegible* but not **in-non-legible*. This fact is reflected in

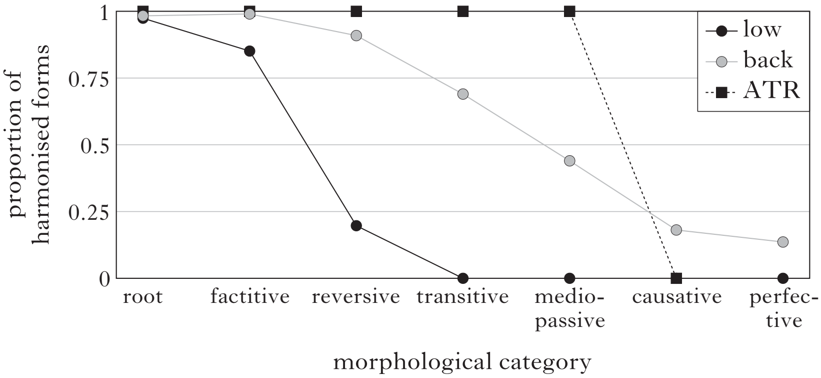


Figure 1
Tommo So vowel harmony: application rates by morphological layer.

the phonology: the process that assimilates /n/ to [l] before [l] is applicable to *in-* (*illegal, illegible*) but not to *non-* (**nol-legible*). The essential correlation is between morphological distance – as reflected in the general affix-ordering principles – and phonological process application. It remains true that linearly, the *non-* of *non-legible* is just as close to the root (adjacent) as the *il-* of *illegible*; the criterion of distance is inferred from examination of the morphology as a whole.

We will see that Tommo So constitutes a far more elaborate case of the correlation of ‘closeness’, as diagnosed by affix order and phonology.

3 Background on Tommo So

3.1 Language and data sources

Tommo So is spoken by about 60,000 people living on the Bandiagara Escarpment in Mali (Hochstetler *et al*. 2004). It is documented in a reference grammar by McPherson (2013). To our knowledge there has been no previous theoretical work on Tommo So vowel harmony; Hantgan & Davis (2012) treat harmony in the related Bondu-so, but the patterning of the latter system is quite different.

The data for this article were gathered by the first author in Mali during a total of 14 months of fieldwork between 2008 and 2012. There were four primary consultants, all from the commune of Tédié; their speech is relatively uniform, and it is reasonable to consider the data as reflecting one single dialect of Tommo So.

There exists no large corpus of Tommo So language material. The data we used were obtained by combing through the entirety of McPherson’s field materials (consisting of an extensive lexicon, example sentences and a variety of narratives, traditional stories and conversations) for words containing any of the suffixes under consideration (§3.3); if the form contained no suffix vowel (due to vowel-hiatus resolution; see (11) below), it was deemed uninformative, and was not included. We mined this corpus (2818 forms) for the statistical generalisations given below. While the corpus is hardly a random sample, there is no reason to expect that the words obtained were biased concerning vowel harmony; most were elicited for other reasons, such as analysis of the morphology or of tone. Stylistically, the corpus is fairly uniform: 78% of forms come from elicited material. The full corpus is available in the supplementary materials.

3.2 The vowel system

The Tommo So vowel inventory is given in (1).

(1) i (ɨ) u iː uː
e o eː oː
ɛ ɔ ɛː ɔː
a aː

There are seven contrastive vowel qualities, [i e ɛ a ɔ o u]; minimal and near-minimal sets are given in (2a), with the corresponding long vowel phonemes illustrated in (2b). Long and short vowels of the same quality behave identically in harmony.[1]

(2) a.	/i/	[bíl]	'ladder'
	/e/	[běl]	'grass'
	/ɛ/	[bɛ̌l]	'animal'
	/a/	[kìdɛ̀ bǎl]	'gathered thing'
	/ɔ/	[àŋà bɔ́l]	'mouth sore'
	/o/	[ból]	'sweep up'
	/u/	[búl]	'smallpox'
b.	/iː/	[gǐːr]	'talisman'
	/eː/	[dʑéːlé]	'goat's waddle'
	/ɛː/	[dʑɛ́ːlɛ̀]	'bring'
	/aː/	[dʑàːlá]	'sweep a little'
	/ɔː/	[dʑɔ̀ːlɔ́]	'rooster's waddle'
	/oː/	[dʑòːló]	'foot chain'
	/uː/	[dʑúːló]	'twin'

Additionally, [i] and [u] are often reduced to the high central vowel [ɨ] in medial position. In this environment, there is no phonemic contrast among the high vowels [i u ɨ]; what one hears is often phonetically intermediate. In faster speech [ɨ] normally appears, whereas in slower speech the

---

[1] Tone is phonemic, contrasting High, Low and toneless; see McPherson (2013: ch. 4). Links to sound files for the examples in (2) can be found in the online version of the journal. All examples are embedded in the sentence frame /__g-ì-m (X say-PERF-1SG) 'I said X', and are repeated three times.

output tends to be closer to [i] or [u]. The latter choice is determined in part by whether the reduced vowel is in a root or a suffix. In roots, the determining factor is normally the preceding vowel ([i e ɛ] tend to favour [i]-like qualities; [u o ɔ] favour [u]-like qualities). When the root vowel is [a], the quality of the reduced vowel is influenced by the place of articulation of neighbouring consonants (labials tend to prefer [u], coronals [i]). In the three suffixes that include a reduced vowel (reversive, transitive and mediopassive; §3.3), the reduced vowel tends to surface in slower speech as [i], though both [ɨ] and coarticulatorily induced [u] are also observed.

Vowel reduction is important phonologically, because reduced vowels are transparent to harmony. The underlying representations of reduced vowels is indeterminate (since their backness and rounding are not contrastive); we will, somewhat arbitrarily, give underlying representations with their most typical surface vowel.

We adopt the feature assignments for Tommo So vowels in (3); ∅ indicates underspecification. The feature [reduced] is employed as an ad hoc stand-in; in a fully formalised theory reduced vowels would be identified by their weak position in metrical structure.

(3) *Features for Tommo So vowels*

	[high]	[low]	[back]	[ATR]	[reduced]
i	+	–	–	∅	–
e	–	–	–	+	–
ɛ	–	–	–	–	–
a	–	+	∅	∅	–
ɔ	–	–	+	–	–
o	–	–	+	+	–
u	+	–	+	∅	–
ɨ	+	–	∅	∅	+

The distribution of the feature [round] is predictable, since vowels are [+round] when [+back] and otherwise [−round]; for brevity we omit the straightforward rules or constraints that would be needed to fill in this value. It is sensible to treat [back] as the phonologically active feature; [a] turns out to be a non-trigger of Backness Harmony (see §4.3 below), and this may plausibly be related to the fact that it is phonetically neither front nor back. We leave [ATR] unspecified in non-mid vowels simply as an expression of agnosticism; we have no phonological or phonetic evidence to justify a classification.

3.3 Verbal morphology

In this section we give an overview of Tommo So verbal morphology, focusing on the suffixes that demonstrate the affiliation of affix ordering and harmony application; see McPherson (2013: chs 11–12) for further detail. For simplicity at this preliminary stage we give examples that happen not to involve vowel harmony.

3.3.1 *Derivational suffixes*. The factitive, which we treat as underlying /-ndɛ́/, derives transitive verbs from intransitive ones (often with causative meaning), as in [dʑímɛ́] 'be injured' ~ [dʑímɛ́-ndɛ́] 'injure'. It can also be used to derive inchoative verbs from adjectives (always in conjunction with the mediopassive suffix), as in [pàlá] 'long' ~ [pálá-nd-íjɛ́] 'become long'.[2]

The reversive suffix is /-ílɛ́/, as in [dɛ̀bɛ́] 'get stuck' ~ [dɛ̀b-ílɛ́] 'get unstuck', while transitive /-íɾɛ́/ denotes that the subject is performing the action of the verb to or on someone else, as in [tɛ́mbɛ́] 'find oneself in a situation' ~ [tɛ́mb-íɾɛ́] 'make somebody find something'. Mediopassive /-íjɛ́/ denotes that the subject is performing the action on herself; thus [pɛ́ːndɛ́] 'spread out (objects)' ~ [pɛ́ːnd-íjɛ́] '(group) spread selves out', and the causative suffix is /-mɔ́/, as in [sɛ́mɛ́] 'slaughter' ~ [sɛ́mɛ́-mɔ́] 'make slaughter'.

3.3.2 *Inflectional suffixes*. Only one inflectional suffix undergoes vowel harmony: the defocalised perfective, which is the version of the perfective employed when some element in the clause other than the verb is focused. This suffix has two allomorphs, /-ì/ and /-è/. Their distribution is somewhat complex, and the harmonic behaviour of the two is somewhat different. We will only discuss the /-ì/ allomorph in this paper, since the /-è/ allomorph never co-occurs with other suffixes, making it impossible to justify a morphological layer on the basis of affix ordering. For further discussion see McPherson (2013: §12.4). An example of the /-ì/ allomorph is [nɔ́ːlɔ́] 'mix' ~ [nɔ̀ːl-ì] 'mixed'.

There are a fair number of other inflectional suffixes, all of which appear outside of the derivational suffixes. None alternates by harmony. They include /-éːlè/ (negative imperfective), /-dɛ̀/ (affirmative imperfective), /-aː/ (perfective non-final), /-eː/ (imperfective non-final), /-lí/ (negative perfective), /-gú/ (negative imperative) and /-mɔ́/ (hortative).

3.4 The system of vowel phonemes in suffixes

We have seen above that a wide variety of contrasting vowel qualities appear in Tommo So verbal suffixes: /i u a e ɛ ɔ/ all occur (we assume the absence of /o/ is an accidental gap). This would likely defeat any effort to derive suffix vowel qualities by some sort of default insertion processes (as a reviewer suggested to us), and in what follows we will assume fully specified underlying suffix vowels.

3.5 Suffix ordering

Here we set out the principles of affix ordering in Tommo So, extending Kiparsky's ordering test to the more elaborate Tommo So ordering pattern. What emerges is a whole chain of pairwise orderings, diagnosing a system of morphological levels or layers.

---

[2] Underlyingly /pálá-ndɛ́-íjɛ́/; for the hiatus resolution seen in this and other examples see §4.2.

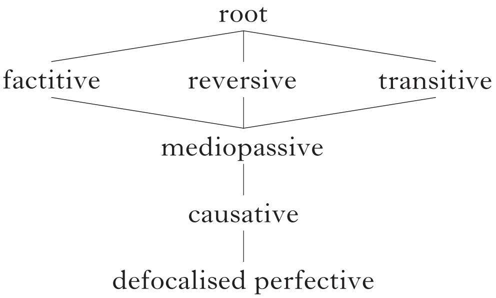
To begin, we find that the factitive precedes the mediopassive, as in [pálá-nd-íjɛ́] (long-FACT-MEDIOPASS) ‘become long’. The opposite order is never found (*[X-íjɛ́-ndɛ́]). The same test with other pairs yields the same result: only one order is possible. For brevity, we summarise the possible combinations in (4).

(4) a. *Factitive before mediopassive*
íɾé-nd-íjé (better-FACT-MEDIOPASS) ‘get better’
b. *Reversive before mediopassive*
mɛ̀nn-íl-íjɛ́ (fold-REVERS-MEDIOPASS) ‘become unfolded’
c. *Transitive before mediopassive*
só-íɾ-íjé (sweat-TRANS-MEDIOPASS) ‘sweat’
d. *Mediopassive before causative*
jùb-íjɛ́-mɔ́ (spill-MEDIOPASS-CAUS) ‘spill’
e. *Causative before perfective*
ɛ̀bɛ̀-m-ì (buy-CAUS-PERF) ‘made buy’

Moreover, the combinations we might expect on grounds of transitivity (like factitive before causative) are generally attested, a point we will not document here. Naturally, in some words there are multiple suffixes present, and these reflect the same ordering generalisations as the two-suffix forms; e.g. [àmà-nd-ìjɛ̀-m-ì] (rancid-FACT-MEDIOPASS-CAUS-PERF).

To test our affix-ordering principles, we used the entire database, extracting from it all words in which two or more verbal affixes attach to the same root. Ordering is entirely consistent: there are no cases whatsoever in which two affixes occur in the opposite order to that given in (4). However, there is one set of affixes (factitive, reversive, transitive) that never co-occur, and thus cannot be assessed for linear order. The results of our ordering study are given in (5) as a Hasse diagram.

(5) *Affix ordering in Tommo So verbs*



For purposes of the analysis to follow, we elaborate these empirical findings into a system of morphological layers, as shown in (6), along the lines envisioned in the theory of Lexical Phonology (Kiparsky 1982, 1994, 2000, 2003).

(6) *Morphological layers assumed for Tommo So*

1. root		
2. factitive	/-ndɛ́/	(derivation)
3. reversive	/-ílɛ́/	(derivation)
4. transitive	/-írɛ́/	(derivation)
5. mediopassive	/-íjɛ́/	(derivation)
6. causative	/-mɔ́/	(derivation)
7. defocalised perfective	/-ì/	(inflection)

These reflect the ordering observations summarised in (5), but go further, in placing factitive, reversive and transitive in separate layers. The placement of transitive in a layer ‘outside’ factitive and reversive can be defended on the grounds that it is more productive and semantically transparent than the latter two suffixes; it is a characteristic of most level-ordered morphological systems that productive and transparent affixes gravitate to outer layers (see e.g. Katamba 2004). The same reasoning might also justify the placement of reversive outside factitive, although the difference in productivity is not as clear in this case.

The system of levels we propose relies on a set of assumptions about how children project the level system from learning data. Our suggestion is that children seek out the *richest* level system compatible with the data (in contradistinction to the ‘level-economising’ approach of classical Lexical Phonology), and that in setting up levels they rely on multiple sources of information. Of these, pairwise affix order is most crucial; that children attend closely to pairwise order is demonstrated by Ryan’s (2010) account of Tagalog. We also assume that children supplement ordering information with further evidence from transparency and productivity. The relatively elaborate structure of (6) is what emerges from this learning process. We suggest that a rich level system is not unique to Tommo So; parallel systems are seen wherever linguists have posited rich ‘position-class’ morphology, as in Athabaskan (e.g. Hargus 1988), Nimboran (Inkelas 1993) and various of the languages given in Nida (1949).

4 The vowel-harmony pattern

With the principles of affix ordering in place, we turn to the other side of the Kiparskian correlation: the applicability of phonological processes, here vowel harmony.

We give the facts of Tommo So vowel harmony first in rule-based phonology (Chomsky & Halle 1968), employed for its descriptive precision. In this framework, Tommo So would be considered to have three vowel-harmony rules, one each for the features [low], [back] and [ATR]. Since harmony is frequently optional, we will give examples for each harmony process of both application and non-application.

4.1 Verb roots

As in other languages (see e.g. Kiparsky 1973: 36 on Finnish and Clements & Sezer 1982: 222–231 on Turkish), the vowel-harmony pattern applies slightly differently within roots and affixes, so we discuss the two separately. Harmony is also slightly different for verbs *vs.* nouns, and we will be focusing on verbs here. For roots, we give data based on a corpus of verbal roots (all of the affixes we discuss are verbal suffixes), which appears as Table I.

		V2, V3						
		i	e	ɛ	a	ɔ	o	u
V1	i	9	39	56	░	▓	▓	▓
	e		37	■	░	■	▓	▓
	ɛ	4	■	79	░	▓	■	▓
	a	2	░	░ 4	151	░	░	4
	ɔ	▓	■	▓ 2	░	100	■	2
	o	▓	▓	■	░	■	46	4
	u	▓	▓	▓	░	49	43	8

Table I

Sequences of vowels in Tommo So verb roots. Disharmonic sequences are shaded (ATR disharmonic = black; Backness disharmonic = dark grey; Low disharmonic = light grey).

The rows of Table I are labelled with the first vowel of a root, the columns with the second or third vowels (roots are maximally three syllables); thus the '9' in the upper leftmost cell indicates nine cases in the corpus in which a root has [i] as its first vowel and [i] as its second or third vowel.

The boxed regions of the table, showing non-initial high vowels, require comment. In Tommo So, verb roots may not end in a high vowel; thus all of these high vowel counts represent *medial* high vowels in trisyllabic roots. As such, they are in the context for vowel reduction, and the observed distribution between [i] and [u] reflects the allophonic variation among reduced vowels noted in §3.2.

It is evident that many logically possible sequences of vowels are absent or severely underrepresented. These gaps are due to vowel harmony. The black cells represent forms excluded by ATR Harmony, the dark grey boxes represent forms excluded (with just two exceptions) by Backness Harmony, and the light grey boxes forms excluded (with just four exceptions) by Low Harmony. We turn now to a detailed description of each harmony process.

4.2 Low Harmony

We state Low Harmony as in (7).[3]

(7) *Low Harmony*

$$\begin{bmatrix}\text{V}\\ -\text{reduced}\end{bmatrix} \rightarrow [\alpha\,\text{low}]\ /\ \text{C}_0 \begin{bmatrix}\text{V}\\ \alpha\,\text{low}\end{bmatrix}\ \text{X}\_\_$$

Here X stands for any sequence, meaning that the rule can apply non-locally, affecting all the non-initial vowels of the word.[4] The surface pattern implied by (7) is as follows. [a] is never followed by a mid vowel ([−high, −low]), nor can it ever appear after an initial vowel other than [a]. Initial high vowels trigger Low Harmony, since in that position they are not reduced (§3.2). Thus, an initial high is never followed by [a]. But high vowels may follow initial [a], either because they are medial and therefore reduced, or because they occur in the defocalised perfective suffix and are therefore outside the domain of Low Harmony; see (12).

The effects of Low Harmony can be seen clearly in our verb-root corpus. Some representative data illustrating the patterns in Table I, as well as exceptions, are given in (8).

(8) *Low Harmony in roots*

a. *Regular forms*

[ámá]	‘be fattened’
[dàmbá]	‘push’
[ádɨ́bá]	‘think’
[dɛ̀nnɛ́]	‘look for’
[súmmɔ́]	‘dilute’

b. *Exceptional forms* (*rare*)

[jàmíɲdʑɛ́]	‘rub soap’
[sáːdɛ́]	‘die without being slaughtered’

The words in (8b) represent two out of the total of only four exceptions in roots.

Consider next the behaviour of Low Harmony in suffixes. Because of ‘petering out’, Low Harmony affects only the two suffixes that form the innermost layers of the morphology in (6). The factitive suffix surfaces as

---

[3] We gloss over a fairly major though orthogonal issue concerning the relationship of harmony to prosodic structure. A Finnish precedent for our claim that initial, prosodically prominent vowels can trigger harmony non-locally is given by Ringen & Heinämäki (1999: 316).

[4] Reviewers ask if harmony is statistically more reliable in local contexts, i.e. when the trigger and target are in adjacent syllables. It appears that there is a modest effect of this kind: notably, disyllabic stems have only about a 1% harmony exception rate, while for the final vowels of trisyllabic stems it is about 7%. In principle we could complicate our analyses to include the local/non-local distinction, but trisyllabic stems are so rare (about 10% of the total) that the pay-off in accuracy would be small.

[-ndá] when it follows a root with an initial low vowel, but as any of [-ndé], [-ndɛ́], [-ndó] or [-ndɔ́] in other contexts, depending on other vowel-harmony processes. /-ndɛ́/ is the underlying value, which surfaces when no harmony process applies.

(9) *Low Harmony in factitive forms*
a. /dʑǎː-ndɛ́/ (meal-FACT) → [dʑàːndá] 'cook'
b. /dàgá-ndɛ́/ (be.good-FACT) → [dàgándá] 'fix'

The application rate of Low Harmony for factitive forms is about 85%. By this we mean that, of all cases in the data where a low-vowel root precedes the factitive suffix, 85% surface with a lowered factitive vowel. An example of non-application is [dʑàːndɛ́], which is the very same word as (9a), uttered by the same speaker on a different occasion.

We treat the reversive suffix as underlying /-ílɛ́/. For this suffix, the application rate of Low Harmony is only 20%. When Low Harmony applies, the reversive surfaces as [-ílá]; otherwise, it appears after [a] as its underlying form [-ílɛ́], or else as [-ílé], [-íló] or [-ílɔ́] where other vowel-harmony processes are applicable. It can be observed that application of Low Harmony to this suffix is non-local, skipping over the reduced suffix-initial vowel.

(10) *Low Harmony in reversive forms*
a. *application*
/màná-ílɛ́/ (seal-REVERS) → [mànálá] 'unseal'
/pándá-ílɛ́/ (widow-REVERS) → [pándílá] 'marry a widow'
b. *non-application*
/jàmbá-ílɛ́/ (cover-REVERS) → [jàmbílɛ́] 'uncover'
/pándá-ílɛ́/ (widow-REVERS) → [pándílɛ́] (free variation with (a))

The forms in (10) illustrate another phonological process of Tommo So, the resolution of hiatus. Normally, when suffixation creates a sequence of two vowels, the first of the two is deleted (McPherson 2013: §3.7). The process is stated in (11).

(11) *Hiatus Resolution*
$V \rightarrow \emptyset \,/\, \_\_\, V$

Hiatus Resolution can be seen in (10), where the final /a/ deletes before suffix-initial /i/ (the second form in (10a) is an irregular case, with hiatus resolved by loss of the second vowel).

The remaining suffixes fall outside the domain in which Low Harmony applies. In (12), we give examples of its non-application for these suffixes.

(12) *Non-application of Low Harmony in forms with the remaining suffixes*

/jàmbá-íɾɛ́/ (cover-TRANS) → [jàmbíɾɛ́] ‘cover’
*[jàmbíɾá]
/jàmbá-íjɛ́/ (cover-MEDIOPASS) → [jàmbíjɛ́] ‘cover oneself’
*[jàmbíjá]
/káná-mɔ́/ (do-CAUS) → [kánámɔ́] ‘make do’
*[kánámá]
/káná-ì/ (do-PERF) → [kànì] ‘did’[5]
*[kànà]

4.2.1 *Excursus: defining application rate.* Since application rates are the key data in this article, we take a moment to define them, using examples from the data just presented. There are four relevant categories, given in (13).

(13) a. *Non-vacuous application*
The process is applicable, and applies, giving a change in the output.
Example: /dʑǎː-ndɛ́/ → [dʑàːndá] (meal-FACT) ‘cook’
b. *Non-application*
The process is applicable, but does not apply, resulting in surface disharmony.
Example: /dʑǎː-ndɛ́/ → [dʑàːndɛ́]
c. *Vacuous application*
The process is applicable, but its conditions were already met in the input, so application is vacuous.
Example: /jɛ̀-ndɛ́/ → [jɛ̀ndɛ́] (see-FACT) ‘look at’
d. *Inapplicability*
The process is not applicable. This never arises for Low Harmony, since it is triggered by any vowel. A legitimate example arises below for Backness Harmony: by (15) below it is not triggered by the low vowel [a], so any disyllabic form whose first vowel is [a] would be a case of inapplicability.

In what follows, we will define application rate by the formula in (14).

(14)
$$\frac{\textit{cases of non-vacuous application}}{\textit{cases of non-vacuous application} + \textit{cases of non-application}}$$

The numerator counts forms in which application produces an observable change, and the denominator counts all cases of potential applicability. This definition is used because it matches the criterion of adequacy for a constraint-based analysis (see §6 below). In cases of vacuous

---

[5] The defocalised perfective form is characterised by an all-L grammatical tone pattern (McPherson 2013: §12.4.1).

application or inapplicability, the faithful candidate incurs neither markedness nor faithfulness violations and will always win, making it uninformative about ranking or weighting.

4.3 Backness Harmony

The Backness Harmony pattern can be described with the rule in (15).

(15) *Backness Harmony*

$$\begin{bmatrix} \text{V} \\ -\text{reduced} \end{bmatrix} \rightarrow [\alpha\text{back}] \;/\; \#\ \text{C}_0 \begin{bmatrix} \text{V} \\ \alpha\,\text{back} \end{bmatrix} \text{X}\,\_\_$$

We first consider some details of the pattern. First, under the feature system we assume, [a] cannot be a trigger of Backness Harmony, since it has no backness value that can be transmitted to the target vowel. Hence [a] is compatible with both front and back following vowels. Moreover, it is impossible to tell if [a] is an undergoer of Backness Harmony, because there are no possible inputs; in roots, Low Harmony eliminates all non-initial [a]'s after backness-specified vowels, and there are no low-vowel suffixes in the layers of the morphology where harmony prevails. Second, Backness Harmony does not target reduced vowels (see §3.2 above), because, as noted above, their backness is gradient and determined by coarticulation.

Turning to the data, we first illustrate Backness Harmony with data from roots, as summarised above in Table I. With very rare exceptions (about 1.7% of all forms), the front vowels [i e ɛ] may only co-occur with front vowels, and the back vowels [u o ɔ] only with back vowels (see the dark grey region of table). In (16) we give examples of both normal harmonic roots and the rare exceptions.

(16) *Backness Harmony in roots*

a. *Regular forms*		b. *Exceptional forms*	
[gìjé]	'harvest'	[gɔ̀bɔ́dɛ́]	'barely touch'
[kɛ́ɾɛ́]	'bite'	[kɔ́jɛ́]	'be hoarse'
[dùgó]	'casts spells'		
[bògóló]	'bellow'		

In suffixes, we find a consistent pattern of 'petering out' as we move morphologically away from the root. We consider the suffixes in order below.

In the factitive, Backness Harmony is virtually exceptionless; (17b) is the only exceptional form in the corpus. For completeness, we include cases of both vacuous and non-vacuous application.

(17) *Backness Harmony in factitive forms*

a. *application*			
/dɔ̌ː-ndɛ́/ (arrive-FACT)	→	[dɔ̀ːndɔ́]	‘move near’[6]
/gǒː-ndɛ́/ (go.out-FACT)	→	[gòːndó]	‘take out’
/dúː-ndɛ́/ (bottom-FACT)	→	[dùːndó]	‘put down’[7]
b. *non-application*			
/dʑɔ̀bɔ́-ndɛ́/ (run-FACT)	→	[dʑɔ̀bɔ́ndɛ́]	‘make run’
c. *vacuous application*			
/jɛ̀-ndɛ́/ (see-FACT)	→	[jɛ̀ndɛ́]	‘look at’
/děː-ndɛ́/ (know-FACT)	→	[dèːndé]	‘introduce’
/dʑímɛ́-ndɛ́/ (be.hurt-FACT)	→	[dʑímɛ́ndɛ́]	‘hurt’
d. *inapplicability*			
/dʑǎː-ndɛ́/ (meal-FACT)	→	[dʑàːndá] *or* [dʑàːndɛ́]	‘cook’

(17d) shows optional application of Low Harmony. It also forms part of the evidence that [a] is not a Backness Harmony trigger: in the variant where Low Harmony does not apply, we find [ɛ] as the suffix vowel, which reflects the underlying form. In all other forms in (17), the suffix also harmonises for the feature [ATR], to be discussed below.

The next morphological layer consists of the reversive suffix. Here Backness Harmony is only slightly less robust, applying to 91% of applicable forms.

(18) *Backness Harmony in reversive forms*

a. *application*			
/gɔ̀ŋɔ́-ílɛ́/ (fence.in-REVERS)	→	[gɔ̀ŋílɔ́]	‘unfence’
/tóɲɲó-ílɛ́/ (crumple-REVERS)	→	[tóɲɲíló]	‘uncrumple’
/mùnnɔ́-ílɛ́/ (roll-REVERS)	→	[mùnnílɔ́]	‘unroll’
b. *non-application*			
/mùɲdʑɔ́-ílɛ́/ (break-REVERS)	→	[mùɲdʑílɛ́]	‘break off’
/úmɔ́-ílɛ́/ (breathe-REVERS)	→	[úmílɛ́]	‘resuscitate’
c. *vacuous application*			
/dɛ̀bɛ́-ílɛ́/ (get.stuck-REVERS)	→	[dɛ̀bílɛ́]	‘get unstuck’
/némbé-ílɛ́/ (trim-REVERS)	→	[némbílé]	‘cut off branch’
/dìŋɛ́-ílɛ́/ (tie-REVERS)	→	[dìŋílɛ́]	‘untie’
d. *inapplicability*			
/tágá-ílɛ́/ (shoe-REVERS)	→	[tágílɛ́]	‘take off shoes’

---

[6] The tonal alternation is due to Tonal Absorption (McPherson 2013: §4.3.1); Rise becomes Low before High.

[7] This form has a noun base; we include it to show that [u] is a backness harmony trigger; for discussion of ATR harmony with noun bases, see note 14.

The vowel transcribed as [i] in these forms is a reduced vowel (see §3.2), and is hence skipped over by harmony. Once again, we see ATR alternations as well as backness alternations.

At the next morphological layer, transitive, the application rate of Backness Harmony drops to 69%. Again, we see that [a] is not a trigger, that the reduced vowel [i] in the suffix is transparent and that backness alternations are accompanied by ATR alternations.

(19) *Backness Harmony in transitive forms*

a. *application*

/ɔ́g-íɾɛ́/ (hot-TRANS) → [ɔ́gíɾɔ́] ‘heat’[8]
/dògó-íɾɛ́/ (be.face.up-TRANS) → [dògíɾó] ‘hold face up’
/túŋɔ́-íɾɛ́/ (kneel-TRANS) → [túŋíɾɔ́] ‘make kneel’

b. *non-application*

/ɔ́g-íɾɛ́/ (hot-TRANS) → [ɔ́gíɾɛ́] ‘heat’[9]
/sónnúgó-íɾɛ́/ (place on shoulders-TRANS) → [úmílɛ́] ‘put on somebody else’s shoulders’

c. *vacuous application*

/sɛ́ː-íɾɛ́/ (adorn-TRANS) → [sɛ́íɾɛ́] ‘adorn’[10]
/tégé-íɾɛ́/ (drip-TRANS) → [tégíɾé] ‘make drip’
/dìmbɛ́-íɾɛ́/ (follow-TRANS) → [dìmbíɾɛ́] ‘make follow’

d. *inapplicability*

/tágá-íɾɛ́/ (shoe-TRANS) → [tágíɾɛ́] ‘put shoes on somebody’

In the mediopassive (20), the application rate of Backness Harmony is 44%.

(20) *Backness Harmony in mediopassive forms*

a. *application*

/tɔ́mɔ́-íjɛ́/ (wind.up-MEDIOPASS) → [tɔ́míjɔ́] ‘be wound up’
/tóɲɲó-íjɛ́/ (crumple-MEDIOPASS) → [tóɲɲíjó] ‘be crumpled’
/jùbɔ́-íjɛ́/ (spill-MEDIOPASS) → [jùbíjɔ́] ‘be spilled’

b. *non-application*

/tɔ́mɔ́-íjɛ́/ (wind.up-MEDIOPASS) → [tɔ́míjɛ́] ‘be wound up’
/góɾó-íjɛ́/ (hat-MEDIOPASS) → [gòɾíjé] ‘wear a hat’
/mùnnɔ́-íjɛ́/ (roll-MEDIOPASS) → [mùnníjɛ́] ‘be rolled up’

---

[8] /ɔ́g/ is an adjectival root. This may be a case of deadjectival derivation; alternatively, we could set up the bound verbal root /ɔ́gɔ́/, whose second vowel is always lost by Hiatus Resolution.

[9] This is the same input form as in (a); indeed, the two free variants were uttered in the same session, about ten minutes apart.

[10] Bound root /sɛ́ː/; long vowels shorten rather than deleting prevocalically.

c. *vacuous application*

/dʑèlé-íjɛ́/ (hang-MEDIOPASS)	→ [dʑèlíjé]	‘be hanging’
/tímbɛ́-íjɛ́/ (stack-MEDIOPASS)	→ [tímbíjɛ́]	‘be stacked’
/pɛ́ndɛ́-íjɛ́/ (make.tight-MEDIOPASS)	→ [pɛ́ndíjɛ́]	‘get crowded’

d. *inapplicability*

/káná-íjɛ́/ (do-MEDIOPASS)	→ [káníjɛ́]	‘take place’

In the causative (21), the application rate is just 18%.

(21) *Backness Harmony in causative forms*

a. *application*

/tɛ́mɛ́-mɔ́/ (eat-CAUS)	→ [tɛ́mɛ́mɛ́]	‘make eat’
/bìɾɛ́-mɔ́/ (work-CAUS)	→ [bìɾɛ́mɛ́]	‘make work’

b. *non-application*

/kɛ́ɾɛ́-mɔ́/ (bite-CAUS)	→ [kɛ́ɾɛ́mɔ́]	‘make bite’
/jèːmé-mɔ́/ (melt-CAUS)	→ [jèːmémɔ́]	‘make melt’
/sídɛ́-mɔ́/ (pay-CAUS)	→ [sídɛ́mɔ́]	‘make pay’

c. *vacuous application*

/dʑɔ̀bɔ́-mɔ́/ (run-CAUS)	→ [dʑɔ̀bɔ́mɔ́]	‘make run’
/óbó-mɔ́/ (give-CAUS)	→ [óbómɔ́]	‘make give’
/nújɔ́-mɔ́/ (sing-CAUS)	→ [nújɔ́mɔ́]	‘make sing’

d. *inapplicability*

/káná-mɔ́/ (do-CAUS)	→ [kánámɔ́]	‘make do’

The form [kánámɔ́] in (21d) should be compared with (20d), [káníjɛ́]: the pair illustrates that [a] is not a Backness Harmony trigger and that the underlying backness value of the suffix surfaces after [a] roots; i.e. /-mɔ́/ *vs.* /-íjɛ́/. Since /-mɔ́/ is underlyingly [+back], our examples of non-application involve front-vowel roots.

Finally, in the defocalised perfective (22), the Backness Harmony rate is 14%.

(22) *Backness Harmony in defocalised perfective forms*

a. *application*

/bòdó-i/ (put-PERF)	→ [bòdù]	‘put’
/óbó-i/ (give-PERF)	→ [òbù]	‘gave’
/dʑùŋgó-i/ (nod-PERF)	→ [dʑùŋgù]	‘nodded’

b. *non-application*

/bɔ̀gɔ́lɔ́-i/ (chatter-PERF)	→ [bɔ̀gɔ̀lì]	‘chattered’
/bòdó-i/ (put-PERF)	→ [bòdì]	‘put’
/dʑùŋgó-i/ (nod-PERF)	→ [dʑùŋgì]	‘nodded’

c. *vacuous application*
/sέgíɾέ-i/ (meet-PERF) → [sὲgìɾì] ‘met’
/dʑ́ídʑíbé-i/ (shake-PERF) → [dʑìdʑìbì] ‘shook’

d. *inapplicability*
/káná-i/ (do-PERF) → [kànì] ‘did’

All remaining morphology has a Backness Harmony rate of zero; by this point, Low and ATR Harmony have also petered out. Thus the suffixes (all inflectional; §3.3) have a single surface realisation.

4.4 ATR Harmony

The feature [ATR] is phonemic in Tommo So only for the class of mid vowels [e ɛ o ɔ]. It is these vowels that form both the trigger and target class for ATR Harmony, as stated in (23).

(23) *ATR Harmony*

$$\begin{bmatrix} \text{V} \\ -\text{high} \\ -\text{low} \end{bmatrix} \rightarrow [\alpha\text{ATR}] \;/\; \begin{bmatrix} \text{V} \\ -\text{high} \\ -\text{low} \\ \alpha\text{ATR} \end{bmatrix} \text{X}\_\_$$

In verb roots, ATR Harmony is exceptionless: see Table I above, where the eight black boxes represent the eight logically possible sequences of disagreeing mid vowels ([e ɛ], [o ɔ], etc.). We give a few representative harmonic roots in (24).

(24) *ATR Harmony in roots*

[έbέ]	‘buy’	[kómmó]	‘crumple’
[gègédé]	‘(insects) nibble’	[gɔ̀bɔ́dέ]	‘barely touch’
[kɔ́ɾɔ́ndɔ́]	‘snore’		

The final form shows that even forms that are exceptions to Backness Harmony obey ATR Harmony.

The high vowels [i] and [u] are not ATR triggers (though see note 11 below): in roots they may co-occur with either the [+ATR] vowels [e o] or the [−ATR] vowels [ɛ ɔ], in either order, as shown in (25).

(25) *Free combination of* [i u] *with both values of* [ATR]

[kíːdέ]	‘discuss’	[údɔ́]	‘build’
[píjé]	‘cry’	[túɲdʑó]	‘slap wet laundry against a stone’

In principle, we could test our claim of non-triggerhood with suffix data, but due to the phonotactic restriction on high vowels in verb roots (§4.1), no actual cases arise: all roots contain at least one non-high vowel, and a high vowel will never be the closest vowel to the suffix. For this reason, the rule is formulated with the closest mid vowel as trigger rather than the initial vowel (as in Low Harmony and Backness Harmony).

The low vowel [a] likewise does not trigger ATR Harmony; after [a], suffixes surface unaltered for [ATR]. Thus the mid allomorph of the defocalised perfective (§3.3), which in general is harmonic, surfaces after [a] with its underlying value of [-è] (e.g. /bàlá-è/ (sweep-PERF) → [bàlè] 'swept'). All remaining mid-vowel suffixes have the underlying value [−ATR], and after [a], this is preserved; see below.

We now exemplify the applicability of ATR Harmony in each layer of the morphology. Here the data are far simpler, because the harmony rate is always either 100% or zero. For the factitive, reversive, transitive and mediopassive, the harmony rate is 100%. The actual allomorph that surfaces depends also on the other forms of harmony.[11]

(26) a. *ATR Harmony in factitive forms*
  i. *application*
    /děː-ndɛ́/ (know-FACT) → [dèːndé] 'introduce'
    /gǒː-ndɛ́/ (go.out-FACT) → [gòːndó] 'take out'
  ii. *vacuous application*
    /jɛ̀-ndɛ́/ (see-FACT) → [jɛ̀ndɛ́] 'look at'
    /dʑɔ̀bɔ́-ndɛ́/ (run-FACT) → [dʑɔ̀bɔ́ndɔ́] 'make run'
  iii. *inapplicability*
    /dʑǎː-ndɛ́/ (meal-FACT) → [dʑàːndá] *or* [dʑàːndɛ́] 'cook'
b. *ATR Harmony in reversive forms*
  i. *application*
    /némbé-ílɛ́/ (trim-REVERS) → [némbílé] 'cut off branch'
    /tóɲɲó-ílɛ́/ (crumple-REVERS) → [tóɲɲíló] 'uncrumple'
  ii. *vacuous application*
    /dɛ̀bɛ́-ílɛ́/ (get.stuck-REVERS) → [dɛ̀bílɛ́] 'get unstuck'
    /tɔ́mɔ́-ílɛ́/ (wind-REVERS) → [tɔ́mílɔ́] 'unwind'
  iii. *inapplicability*
    /jàmbá-ílɛ́/ (cover-REVERS) → [jàmbílɛ́] 'uncover'

---

[11] There is a minor conundrum involving the factitive suffix that we have not resolved: the very small amount of evidence (four roots) that we have indicates that high-vowel roots trigger [+ATR] in the factitive; for instance [dùː-ndó] (bottom-FACT) 'put down'. All four roots are nominal or adjectival; as already noted, verb roots cannot end in high vowels, which is why these examples are rare. The appearance of [+ATR] vowels in post-high factitives is a puzzle, since high vowels are fully compatible with [−ATR] vowels in roots (see (25)). To resolve this, we might add an additional minor phonological process specific to the factitive, or perhaps adopt a special underlying representation (e.g. underspecified for [ATR], or multiple listed allomorphs). As this behaviour does not appear to bear on our main point, we will not attempt to resolve it here.

c. *ATR Harmony in transitive forms*
  i. *application*
    /tégé-íɾɛ́/ (drip-TRANS) → [tégíɾé] ‘make drip’
    /dògó-íɾɛ́/ (be.face.up-TRANS) → [dògíɾó] ‘hold face up’
  ii. *vacuous application*
    /tɛ́mbɛ́-íɾɛ́/ (find-TRANS) → [tɛ́mbíɾɛ́] ‘make find’
    /ɔ́g-íɾɛ́/ (hot-TRANS) → [ɔ́gíɾɔ́] ‘heat’
  iii. *inapplicability*
    /tágá-íɾɛ́/ (shoe-TRANS) → [tágíɾɛ́] ‘put shoes on somebody’

d. *ATR Harmony in mediopassive forms*
  i. *application*
    /dʑèlé-íjɛ́/ (hang-MEDIOPASS) → [dʑèlíjé] ‘be hanging’
    /tóɲɲó-íjɛ́/ (crumple-MEDIOPASS) → [tóɲɲíjó] ‘be crumpled’
  ii. *vacuous application*
    /tɔ́mɔ́-íjɛ́/ (wind.up-MEDIOPASS) → [tɔ́míjɛ́] ‘be wound up’
    /pɛ́ːndɛ́-íjɛ́/ (spread out-MEDIOPASS) → [pɛ́ːndíjɛ́] ‘spread out’
  iii. *inapplicability*
    /tágá-íjɛ́/ (shoe-MEDIOPASS) → [tágíjɛ́] ‘put shoes on oneself’

e. *ATR Harmony in causative forms*
  i. *non-application*
    /dʑèɲɲé-mɔ́/ (pick.up-CAUS) → [dʑèɲɲémɔ́] ‘make pick up’
    /óbó-mɔ́/ (give-CAUS) → [óbómɔ́] ‘make give’
  ii. *vacuous application*
    /nɔ́ː-mɔ́/ (drink-CAUS) → [nɔ́ːmɔ́] ‘make drink’
    /ŋ̩́jɛ́-mɔ́/ (eat-CAUS) → [ŋ̩́jɛ́mɔ́] ‘make eat’
  iii. *inapplicablity*
    /wàlá-mɔ́/ (farm-CAUS) → [wàlámɔ́] ‘make farm’

For the causative suffix, the harmony rate is zero. Hence it always surfaces as [−ATR]; usually [-mɔ] but sporadically [-mɛ], due to Backness Harmony.

4.5 Summary of suffix harmony

In Table II we summarise what has been said so far about suffix harmony.[12] The second column gives our proposed underlying representation for each

---

[12] Neither [-ndé] nor [-ílé] occurs after [o] roots in the data corpus. We treat these cases as accidental gaps, resulting from (a) [o] being somewhat less common than other vowels in Tommo So, and (b) Backness Harmony being almost obligatory for the factitive and reversive suffixes. The probability that a total of zero tokens for these allomorphs could arise by chance can be calculated at 0.51 for [-ndé] and 0.2 for [-ílé]; thus there is little support for the view that these absences are meaningful.

suffix	underlying	after [e]	after [ɛ]	after [a]	after [ɔ]	after [o]
factitive	/-ndɛ́/	-ndé$_{\mathrm{AH}}$	-ndɛ́	-ndá$_{\mathrm{LH}}$ -ndɛ́	-ndɔ́$_{\mathrm{BH}}$ -ndɛ́	-ndó$_{\mathrm{BH,AH}}$ *-ndé*$_{\mathrm{AH}}$
reversive	/-ílɛ́/	-ílé$_{\mathrm{AH}}$	-ílɛ́	-ílɛ́ -ílá$_{\mathrm{LH}}$	-ílɔ́$_{\mathrm{BH}}$ -ílɛ́	-íló$_{\mathrm{BH,AH}}$ *-ílé*$_{\mathrm{AH}}$
transitive	/-íɾɛ́/	-íɾé$_{\mathrm{AH}}$	-íɾɛ́	-íɾɛ́	-íɾɔ́$_{\mathrm{BH}}$ -íɾɛ́	-íɾó$_{\mathrm{BH,AH}}$ -íɾé$_{\mathrm{AH}}$
mediopassive	/-íjɛ́/	-íjé$_{\mathrm{AH}}$	-íjɛ́	-íjɛ́	-íjɛ́ -íjɔ́$_{\mathrm{BH}}$	-íjé$_{\mathrm{AH}}$ -íjó$_{\mathrm{BH,AH}}$
causative	/-mɔ́/	-mɔ́ -mɛ́$_{\mathrm{BH}}$	-mɔ́ -mɛ́$_{\mathrm{BH}}$	-mɔ́	-mɔ́	-mɔ́
defocalised perfective	/-ì/	-ì	-ì	-ì	-ì -u$_{\mathrm{BH}}$	-ì -u$_{\mathrm{BH}}$

Table II

Suffix allomorphs.

suffix, and the remaining columns list the surface allomorphs as they appear after the five types of verb root (recall that due to verbal phonotactics there are no verb roots with all high vowels). All surface forms assumed to have undergone some harmony process non-vacuously are labelled. In each cell with more than one allomorph, the more frequent one is listed first.

The factitive and reversive suffixes show a five-way alternation ([e ~ ɛ ~ a ~ ɔ ~ o]), since all three harmony processes apply to them. The transitive and mediopassive suffixes have only four allomorphs ([e ~ ɛ ~ ɔ ~ o]), since Low Harmony peters out before reaching them. The causative has just two allomorphs ([ɛ ~ ɔ]), determined by Backness Harmony, since ATR Harmony peters out before reaching it. The defocalised perfective likewise has only two allomorphs ([i ~ u]), determined by the only applicable harmony process, Backness Harmony.

4.6 Frequency of application

In Table III we compile the application rates reported above for all three harmony processes in all morphological contexts.[13] The frequency values represent token frequency, calculated by counting each instance of a repeated underlying form as a separate case. Moreover, where a form has

---

[13] Root rates are calculated as follows: the denominator is the number of forms that match the structural description of the rules in §4, and the numerator the number of harmony-compliant forms; we make no claims about the underlying representations of harmonised roots. Our method of calculating root rates is somewhat arbitrary, but we assume that other choices would yield very similar results; by any criterion all three harmony processes are very close to being obligatory in roots.

	Low harmony		Backness harmony		ATR harmony	
root	151 / 155	97·4%	470 / 478	98·3%	264 / 264	100%
factitive	57 / 67	85·1%	95 / 96	99·0%	80 / 80	100%
reversive	12 / 61	19·7%	40 / 44	90·9%	43 / 43	100%
transitive	0 / 15	0%	40 / 58	69·0%	31 / 31	100%
mediopassive	0 / 167	0%	107 / 243	44·0%	231 / 231	100%
causative	0 / 42	0%	13 / 72	18·1%	0 / 43	100%
perfective	0 / 119	0%	17 / 125	13·6%	*n/a*	0%

Table III

Application rates of vowel harmony by suffix and harmony process.

two harmonising suffixes (none has three), it is counted twice, once for each suffix; this makes sense, because what we seek to model is the rate at which each suffix undergoes harmony. Application rates are also shown in Fig. 2, fleshing out Fig. 1; error bars represent 95% Clopper–Pearson binomial confidence intervals. As can be seen, for two of the harmony processes, application peters out gradually going outward from the root. ATR Harmony instead plummets from 100% to 0% at the mediopassive–causative break.

The two gradient processes peter out in a particular way, forming sigmoid (S-shaped) curves. When we turn to our analysis below (§6), our primary goal will be to derive this sigmoid shape in terms of principles which are well motivated in linguistic theory.

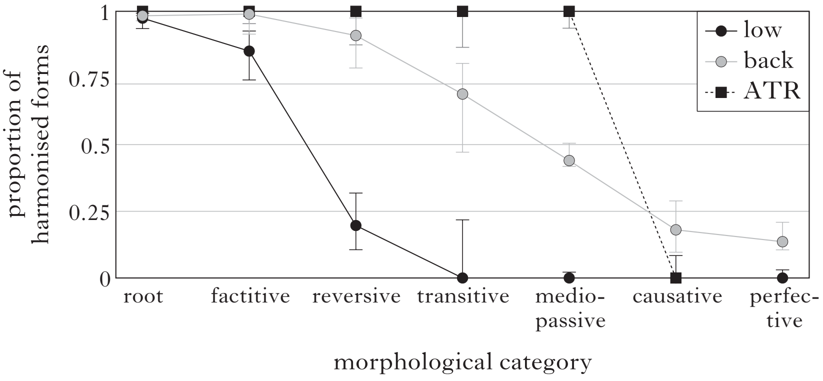


Figure 2

Tommo So vowel harmony: application rates by suffix and harmony type.

5 An earlier proposal: the inverted-exponential model

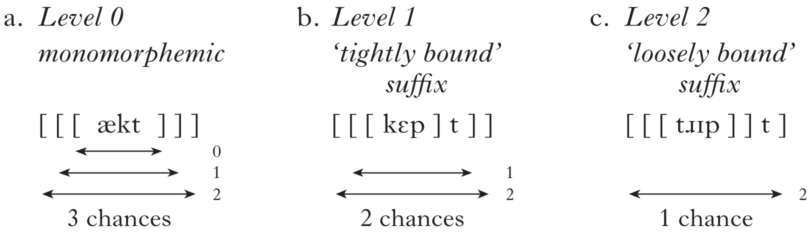
Before doing this, we first review what might be considered the ‘standard model’ for relating morphological distance to phonological process application, namely the widely cited proposal of Guy (1991). Guy developed his model in the course of studying simplification in English final C + {t,d} clusters, which likewise applies with lower frequency in outer layers of the morphology.

In Guy’s model, the diminishing frequency of application emerges directly from the architecture of the grammar. He adopts the general approach to morphological layering assumed in the theory of Lexical Phonology and Morphology, in which layers are the result of a sequence of derivationally ordered levels. For English, the innermost level, Level 0, is the root; thus when the /t/ of a simplex form like *act* [ækt] is dropped in fluent speech, this is a case of a Level 0 /t/. The intermediate Level 1 is the domain of the /t/ that occurs as a non-productive past tense suffix, triggering stem allomorphy in words like *kept* [kɛpt]. The outermost Level 2 is the domain of the regular past tense allomorph seen in words like *tripped* [tɹɪpt].

Studies examining the dropping of alveolar stops in consonant clusters have consistently found a petering out effect (Fasold 1972, Guy 1980, Neu 1980, Nesbitt 1984). For example, in Santa Ana’s study of Chicano English (1991), root-level alveolar stops are deleted 74.3% of the time, intermediate-level stops 59.3% and outermost-level stops 42.1%.

In Guy’s model, such frequencies emerge from the system of level organisation. He proposes that the *-t,d*-deletion rule has a constant application rate, which can vary from dialect to dialect. Moreover, *-t,d*-deletion applies at all three morphological levels, and failure to apply on one level does not preclude application on a later level. Thus, for roots like *act* there are three chances for *-t,d*-deletion to apply, for *kept* there are two and for *tripped* there is just one, as shown in (27).

(27)



Assuming an application rate r, then after n chances to apply, the ‘survival rate’ is $(1 - r)^n$ and the application rate is therefore predicted to be $1-(1-r)^n$. If we plot application rate against n for various values of r, we get a family of INVERTED EXPONENTIAL CURVES, as in Fig. 3.

When a phonological process applies at more than one level, we can predict the application rate for all later levels using the inverted

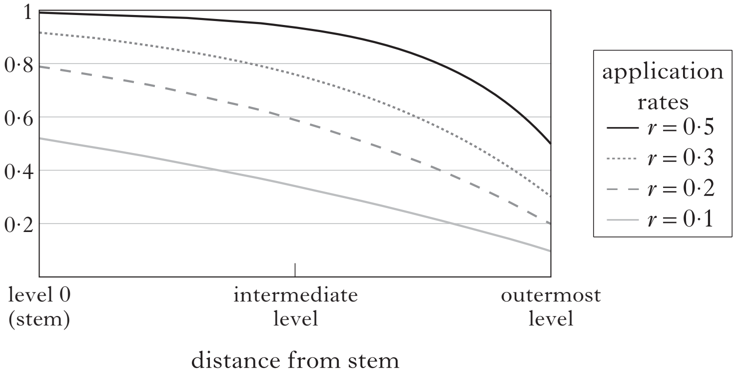


Figure 3

The inverted-exponential model: schematic predictions of application rate.

exponential formula (in the English case, given just one parameter, we predict three observed values). For the data from Santa Ana mentioned above, this is a fairly good fit, as shown in Fig. 4.[14]

We applied the inverted-exponential model to the Tommo So data in Table III. Since we assume seven morphological levels for Tommo So, in the crucial formula $1-(1-r)^n$, n will range from 1 to 7. Using the default settings of the Solver package in Excel, we found for each harmony process the value for r that best fits the data, assuming mean absolute error as our criterion of model fit.[15] We set the application rate so as to minimise model error rather than by selecting any particular level (such as the stem) as criterial; our procedure gives the benefit of the doubt to the model. The resulting fit was poor, as Fig. 5 indicates. Visually, the match between the model and the data is poor; the mean absolute error is 0.181 (error by harmony type: Low 0.249, Backness 0.134, ATR 0.157). Qualitatively, the inverted exponential curves are a poor match to the empirical curves, which, as Fig. 2 above indicates, are not inverted U's, but are S-shaped.

---

[14] For a recent critique of the model as applied to English *-t,d*-deletion, see Fruehwald (2012: 79–80), who suggests that the best-fit values of n in the model formula actually do not come out cleanly as integers when we include a random effect for word identity. For a case where the effect of morphological level failed to reach significance for deletion data, see Tagliamonte & Temple (2005).

[15] The details of model-fitting apparently matter rather little. We rechecked our work with other software and by hand, and also using mean squared error rather than absolute error as our criterion; the values of r that emerged were similar in all cases. The spreadsheets for model fitting are available in the supplementary materials.

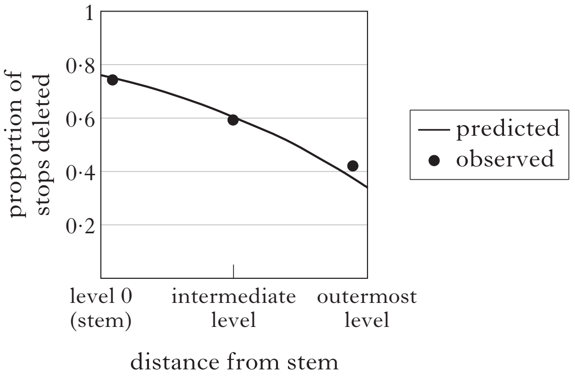


Figure 4

The inverted-exponential model applied to cluster simplication.

6 Deriving the data pattern with Harmonic Grammar

We attempt here to find a better analysis, using ideas from current phonological theory. Ideally, sigmoid shapes will not be stipulated, but will follow naturally from deeper principles.

We adopt two tools from current theorising. First, we employ the constraint-based framework of Harmonic Grammar (Legendre *et al.* 1990, Smolensky & Legendre 2006, Pater 2009a, Potts *et al.* 2010, Jesney 2011, Jesney & Tessier 2011). This framework resembles Optimality Theory (Prince & Smolensky 1993), but uses weighted instead of ranked constraints. In brief, every constraint bears a weight (a real number, representing its strength); every candidate is assigned a value (its HARMONY), which is a sort of penalty score, computed by multiplying weights by

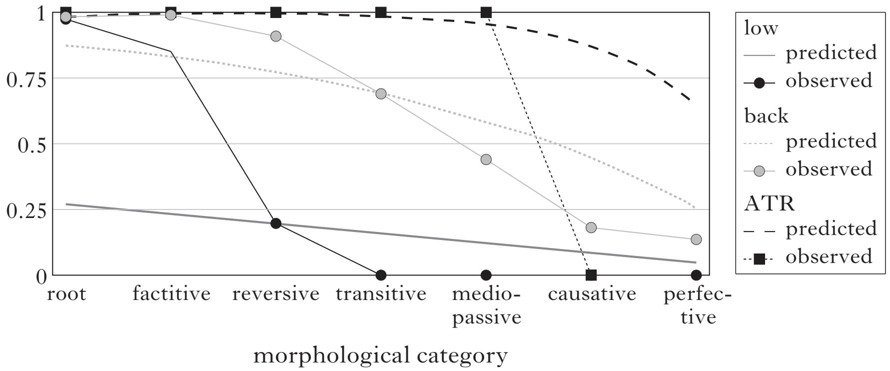


Figure 5

The inverted-exponential model applied to Tommo So harmony.

violation counts and summing. In the simplest version of the theory, the winning candidate is the one with the best (lowest) harmony value.

Since the data involve variation, we need a probabilistic implementation of Harmonic Grammar. There are two such implementations, which turn out to work about equally well. We give first an analysis with maxent grammar (Smolensky 1986, Goldwater & Johnson 2003, Wilson 2006, Hayes & Wilson 2008), then an analysis with Noisy Harmonic Grammar (Boersma & Pater 2016).

The second theoretical tool we will use is the idea that harmony for some constraints is calculated by multiplying the constraint weight by a value along a scale. For us, the scale will be an abstract morphological notion, 'closeness to the root', based on the evidence given in §3.5. The values we adopt for our scale are as in (28).

(28) *The 'root-closeness' scale*

root ≫ factitive ≫ reversive ≫ transitive ≫ mediopassive ≫ causative ≫ perfective
7 ≫ 6 ≫ 5 ≫ 4 ≫ 3 ≫ 2 ≫ 1

The use of scales in constraint-based grammar has a long history. The original work in Optimality Theory, Prince & Smolensky (1993: §5.2), suggested a constraint Hnuc that was gradiently violable, based on the sonority of the segment occupying the nuclear position in a syllable.[16] Frisch *et al*. (1997) and Frisch *et al*. (2004) took the next step, adopting constraint weights one of which was multiplied by the value on the relevant scale (a similarity metric for pairs of consonants in Arabic verbal roots); indeed, the mathematics they employed is a special case of the maxent framework we will be employing. For Flemming (2001), the crucial scale was phonetic, namely values of second-formant frequencies and their (squared) deviations from a target. Using a framework essentially identical to Harmonic Grammar with scalar constraints, Flemming found that he could derive the well-known 'locus effect' in phonetics (Sussman *et al*. 1993) from first principles. A more recent Harmonic Grammar study, Kimper (2011), used scales based on locality, vowel quality and similarity to govern variable application of vowel harmony. We pursue the ideas in these works, using our morphological-distance scale.[17]

To obtain the scale in (28), we arrange the affixes in descending order of root-closeness. Each suffix is separated from its neighbour(s) by the

---

[16] Prince & Smolensky avoid the use of actual integer values in their constraint, but the analytic effect is quite similar.

[17] Our morphological distance scale reduces morphological structure to a single number, as given in (28). In doing this, we do not intend to reject the insights of stratal theories of phonology (Kiparsky 1982, 2000, 2003, Bermúdez-Otero 1999); rather, we assume that root-closeness is an abstract property that not only serves to define the system of strata, but is also directly accessible to constraints; that is, the constraints 'know' what layer a suffix belongs to and can access this information as a scalar value.

arbitrarily chosen value of 1. As we will show, selecting a separation value different from 1 yields exactly the same predictions, once the weights have been fitted to the data. It also makes no difference if we let the scale run between different termini, such 6 to 0 or 0 to −6. Spacing the suffixes along the scale unevenly would change the quantitative predictions of the analysis to a degree which depends on the differences in spacing.[18]

We seek a grammar that, for each combination of morphological layer and harmony process, yields a number expressing the probability of harmony as applied to the relevant suffix. To do this, we adopt a set of ordinary markedness and faithfulness constraints, letting the markedness constraints refer to the root-closeness scale of (28). The constraints will be weighted in a way that generates a close match to the data.

6.1 Maxent calculations

In this section we briefly review the calculations of a maxent grammar; for fuller discussion see for example Goldwater & Johnson (2003) and Hayes & Wilson (2008).

A maxent grammar assigns probabilities to candidates as follows. For each candidate for a given input, the first step is to calculate its harmony by multiplying each constraint weight by the number of violations of that constraint, then summing up across constraints, as in (29).

(29) *Calculation of harmony*

$$H(x) \sum_{i=1}^{n} w_i \, C_i(x)$$

where x is some candidate

$H(x)$ is the harmony value being computed for that candidate

w_i is the weight of the ith constraint

$C_i(x)$ is the number of times that x violates the ith constraint

$\sum_{i=1}^{n}$ denotes summation over all constraints (C_1, C_1, … C_n)

For the present application, the calculation can be made particularly simple if we adopt the idealising assumption that there are only two candidates to consider for each input, one for each value of the harmonising feature. Thus it is implicit in our account that other, non-stated constraints rule out any candidates not mentioned. For example, *[-ndí] for factitive /-ndɛ́/ can be assigned an arbitrarily low probability by giving Ident [high] an indefinitely high weight.

Once the harmony of the two relevant candidates (Cand_1 and Cand_2) has been calculated, the probability of Cand_1 is determined as in (30).

---

[18] For non-linear scales in Harmonic Grammar, see Pater (2016).

(30) *Probability of a candidate in maxent* (*two-candidate system*)

$$p(\text{Cand}_1) = \frac{exp(-H(\text{Cand}_1))}{exp(-H(\text{Cand}_1)) + exp(-H(\text{Cand}_2))}$$

where $p(x)$ is the probability of candidate x
$exp(y)$ is e^y, where e is the base of natural logarithms
$H(x)$ is the harmony of x, from (29)

The probability of Cand_2 is 1 minus the probability of Cand_1. The overall pattern is that large harmony values (deriving from violations of strong constraints) produce low predicted probability values. The probability predicted for a candidate depends on the aggregate strength of its constraint violations, as well as those of the candidate with which it competes.

6.2 Constraint set

In our analysis of Tommo So, harmony is favoured by markedness constraints of the AGREE family (Lombardi 1999, Baković 2000); these are opposed by corresponding faithfulness constraints of the IDENT family (McCarthy & Prince 1995).[19] The three AGREE constraints are given in (31); these are approximations, whose content is to be modified.

(31) a. AGREE[low]
Assign a violation for every non-high vowel that disagrees in [low] with the initial vowel.
b. AGREE[back]
Assign a violation for every vowel that disagrees in [back] with the initial vowel.
c. AGREE[ATR]
Assign a violation for every mid vowel that disagrees in [ATR] with a preceding mid vowel.

The corresponding IDENT constraints are given in (32).

---

[19] The current theoretical climate for the analysis of vowel harmony is unsettled; a wide variety of constraints for enforcing harmony have been put forth. In addition to AGREE, the literature includes two types of ALIGN (of featural autosegments (Kirchner 1993, Ringen & Vago 1998) and of feature domains (Cole & Kisseberth 1994, McCarthy 2004)), as well as SPREAD (Kaun 1995, Walker 1998) and systems based on agreement by correspondence (Rose & Walker 2004, Rhodes 2012). Our concerns are, we think, largely orthogonal to the current debates; what we need is *some* sort of constraint that can enforce featural agreement, and we adopt AGREE as one option from among many.

(32) a. Ident[low]
Assign a violation for every segment that disagrees in [low] with its underlying value.
b. Ident[back]
Assign a violation for every segment that disagrees in [back] with its underlying value.
c. Ident[ATR]
Assign a violation for every segment that disagrees in [ATR] with its underlying value.

6.3 Scalar constraints

The next step is to modify our Agree constraints so that they are sensitive to the morphologically defined scale in (28). The idea is that when the target disagrees with the trigger vowel in the relevant way, the degree of the violation is determined by the morphological closeness of the target. In particular, following the references cited above, we suggest that the number of violations for any particular Agree constraint is simply the value of the scale for the target vowel. In essence, this turns the calculations of maxent from a multiplication of two values (weights × violations) to a multiplication of three (weights × violations × scalar value), as shown in (33).

(33) *Calculation of harmony with scalar constraints*

$$H(x) \sum_{i=1}^{n} w_i \; C_i(x) \; S_i(x)$$

where $S_i(x)$ is the value of candidate x along the scale invoked by constraint C_i

The constraints that invoke a scale in our analysis are the Agree constraints of (31); it should now be understood that for each of these constraints, we are scaling the violations according to (28).

6.4 Establishing the weights

As with other stochastic grammar theories, maxent comes with a learning model. It is assumed that the human language learner brings the constraints to the task of acquisition and that the rankings or weights are determined algorithmically during the learner's encounter with language data (see e.g. Tesar & Smolensky 2000, Boersma & Hayes 2001 and Boersma & Pater 2016). We adopt this approach here, using our data corpus as an approximation of real-life learning data.

For maxent grammar, it matters little in practical terms what optimisation algorithm is adopted for purposes of learning; the search space is free of local maxima (Della Pietra *et al*. 1997), and many procedures yield essentially the same result. We chose to set the weights to minimise mean absolute error, since this is how we report model fit. For convenience and easy replicability we used the Excel Solver application in its default settings. As a check we recomputed the results with the maxent grammar tool (Wilson & George 2009), and obtained very similar results. All weight-setting calculations are posted in the supplementary materials.

We fitted six numbers: the constraint weights for the three Ident constraints and the three Agree constraints. The procedure is non-circular, since we are using six values to predict 20 observations, i.e. the frequency of application for all possible combinations of harmony processes in seven morphological contexts. The weights obtained were as in (34).

(34)	Ident[low]	15.2	Agree[low]	2.8
	Ident[back]	4.0	Agree[back]	1.2
	Ident[ATR]	85.6	Agree[ATR]	34.8

6.5 Results

From the grammar thus constructed we computed the predicted percentage of vowel harmony, using the root-closeness values of (28) and the maxent formulae in (30) and (33). These percentages, given in Table IV, closely fit the original data; in fact, the mean absolute error of the maxent model is 0.012 (broken down by harmony type, i.e. Low 0.012, Backness 0.022, ATR 0.000). This compares very favourably with the 0.181 (Low 0.249, Backness 0.134, ATR 0.157) obtained with the inverted exponential model.

While the superior fit of the model is encouraging, it is also important to understand the model from the viewpoint of restrictiveness – what sorts of frequency patterns could it model *in principle*? For this purpose it helps to consider in general terms how the model makes its predictions.

6.6 Sigmoid curves and their derivation

Harmonic grammars – in the general case – can be quite complex. Prince (1997) showed that under suitable idealising assumptions, a harmonic grammar can mimic the behaviour of any grammar expressed in Optimality Theory, hence all the intricate data patterns that can be derived with ranking in OT can also be derived in maxent. However, in the present case we have made an additional assumption that greatly simplifies the analysis, namely that all candidates other than simple vowel harmony or the faithful realisation are ruled out by other constraints. Hence only two candidates have any chance of winning, and it is solely the weights of Agree and Ident that determine the outcome.

Simple algebra tells us that under these circumstances, the only frequency patterns that can be derived by the grammar will take the form of sigmoid curves, of the particular type known as the logistic function, widely used in statistical analysis.[20] The form of the logistic

---

[20] We suppress the mathematics here (it is textbook material, not research), but invite readers to peruse the supplementary materials.

	Low harmony		Backness harmony		ATR harmony	
	observed	predicted	observed	predicted	observed	predicted
root	97.4	98.4	98.3	99.1	100	100
factitive	85.1	79.5	99.0	97.0	100	100
reversive	19.7	19.7	90.9	90.4	100	100
transitive	0	1.5	69.0	73.1	100	100
mediopassive	0	0.1	44.0	44.0	100	100
causative	0	0	18.1	18.5	0	0
perfective	0	0	13.6	6.2	*n/a*	*n/a*

Table IV

Comparison between the observed application rates of harmony (percentages) and the predictions of the maxent grammar. As a high-voweled suffix, the perfective is ineligible for ATR Harmony. If a mid-voweled suffix occurred in this cell, the application rate predicted by the grammar would be zero.

function relating application frequency to morphological distance is given in (35) (it is obtained by substituting (29) and (33) into (30)). In the formula, we generalise from the specific markedness and faithfulness constraints used here to the case of any phonological process governed by a conflicting scalar markedness constraint and non-scalar faithfulness constraint.

(35) *Sigmoids generated under our analysis*

$$p(\text{candidate undergoing phonology}) = \frac{1}{1 + e^{w_{\text{Faith}} - (w_{\text{Markedness}} \times s)}}$$

where s is the scale value of the candidate

Plugging in our particular weights for markedness and faithfulness in (34) into this formula, we can restate Table IV as Fig. 6, plotting the predictions of the analysis as the sigmoid curves generated by the formula. Further mathematics, given in the supplementary materials, demonstrates the restrictiveness of the approach and illuminates the role of the constraint weights in the analysis. The following mathematical results concerning the sigmoid curves are relevant.

Asymptotes at zero and one. Assuming sufficient room on the horizontal scale, the predicted application rate will asymptote at zero on the low end and one on the high end. Of course, it is possible that the actual values available in the language will not permit us to see these asymptotes; the curves can be ‘cut off’ at one or both ends.

Location of 50% point. The application rate of the phonological process reaches 50% at a point that is determined by the ratio of the constraint weights, specifically at $s = w_{\text{Faith}} / w_{\text{Markedness}}$. For instance, the application

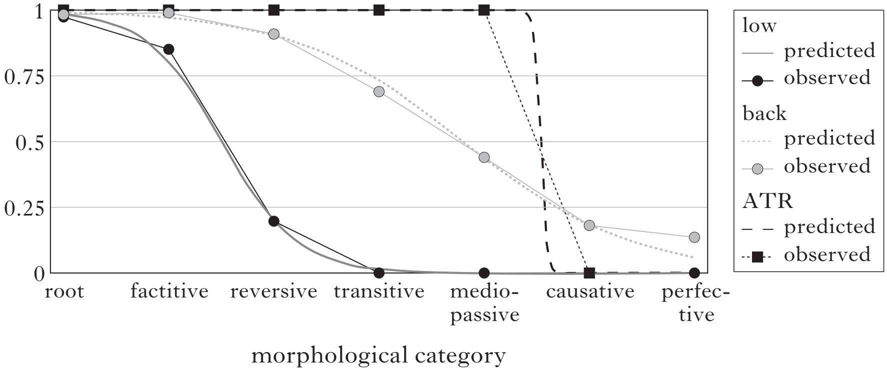


Figure 6

Predictions of the maxent grammar plotted as sigmoid curves.

rate for Low Harmony crosses 50% on our morphological closeness scale at 5.4 (= 15.2/2.8; for weights see (34)).

Symmetry. The sigmoid curve is symmetrical about the 50% point.

Steepest slope. The steepest slope of the curve occurs at the symmetry point and is determined solely by the weight of the markedness constraint; specifically, it is equal to $w_{\text{Markedness}}/4$. For instance, in our analysis the essentially categorical cut-off for ATR Harmony is captured with a very high weight (34.8) for Agree[ATR].

We emphasise that these predictions are specific to the Harmonic Grammar model; other theories do not make these predictions, and for them the Tommo So data are correspondingly problematic; either the model fit is bad, as with the inverted exponentials discussed in §5, or the theory is not restrictive enough and can fit essentially any data (see §7).

We can now justify an assertion made earlier in connection with our root-closeness scale (28): it does not matter what we pick as the baseline value or interval size. As (35) suggests, different choices made for the root-closeness scale are cancelled out when the best-fit weights are found. Thus, if we made the interval size of our scale 2 instead of 1, the best-fit model would use a value for w_{Agree} that was half as big. More generally, the expression $w_{\text{Faith}} - w_{\text{Markednesss}}$ in (35) represents a linear rescaling (y = mx + b) of s, meaning that appropriate choices of weights exist that can compensate for any linear rescaling of (28).

6.7 A variant with Noisy Harmonic Grammar

We have expressed our account in maxent, but this is not the only framework for stochastic analysis in Harmonic Grammar. In Noisy Harmonic Grammar (Boersma & Pater 2016), a 'noise' factor drawn from the Gaussian distribution is added to the weight of each constraint every time the grammar is applied, and the output for any given input for that particular application is simply the most harmonic candidate. The probabilities of the candidates in the general case are defined as the probability distribution over multiple applications. The best-fit constraint weights can be calculated in various ways. Mathematically, Noisy Harmonic Grammar generates symmetrical zero-to-one sigmoid curves, just like maxent, but the sigmoids are from a different function (the cumulative normal distribution), which is strikingly similar visually to the logistic function employed by maxent.

Noisy Harmonic Grammar turned out to work slightly better than maxent, with a mean absolute error of 0.009 (cf. maxent 0.012). Errors for each harmony process are: Low 0.005, Backness 0.021, ATR 0.000.

A technical issue in Noisy Harmonic Grammar is how the noise should be added in the case of scalar constraints: in principle it could be added *before* the harmony contribution is multiplied by the scale value, so that noise itself is multiplied. This produces asymmetrical sigmoids and indeed a slightly worse model fit (mean absolute error 0.031). The value 0.009 just mentioned is obtained by adding the noise *after* multiplication.[21]

6.8 Analysing the English *-t,d*-deletion data in maxent

It is appropriate to ask whether the English *-t,d*-deletion data that motivated Guy's inverted exponential theory can be modelled under our theory. The answer is that it can. In brief, we set up a *-t,d*-deletion model employing the weighted constraints *CT (which penalises clusters; w = 0.69) and MAX(T) (faithfulness; w = 1.01), with the root-closeness scale 3 = root, 2 = Level 1, 1 = Level 2. For Santa Ana's (1991) data this gives highly accurate results.[22]

---

[21] The Noisy Harmonic Grammar analysis can be replicated in the framework of Stochastic OT, following a strategy of Boersma (1998: §6, §8.4). Instead of a single markedness constraint, a family of markedness constraints is adopted, one for each morphological level, spaced at equal intervals along the ranking scale. This, too, generates sigmoid curves from the cumulative normal distribution. The slope of the sigmoid depends on the spacing of the members of the markedness family, and the 50% crossing point is determined by the relative ranking of markedness and faithfulness. The consequences of fragmented constraints of this kind have been little explored in the literature, and the issue of the learnability of ranking values for Stochastic OT is also problematic (Pater 2008).

[22] Our own model has a greater number of parameters and might be expected to have a better fit *a priori*. However, the fit of the inverted-exponential model to Tommo So does not improve substantially when we increase its number of parameters in the most obvious way, by adding an intercept term (which can raise or lower the overall height of the curve). We tried this approach, and found that the mean absolute errors remain high: overall 0.172, Low 0.223, Backness 0.134, ATR 0.156 (earlier 0.181, 0.249, 0.134, 0.157). A reviewer also asked whether our own model could do as well if we reduced the number of parameters. We tried this first by setting markedness to a UG-determined constant (fitting it to data from all three harmony processes at once) then by doing the same for faithfulness. The resulting error values are high; 0.076, 0.068, 0.042, 0.107 for constant markedness, and 0.090, 0.123, 0.055, 0.080 for constant faithfulness.

6.9 Sigmoid curves in linguistic theory

Sigmoids have provided good data fit in various places in linguistics whenever a constraint with scalar violations stacks up against a constraint with constant violations. Albright (2012), in a study of phonotactics, finds that variable phonotactic disharmony scores get a closer fit to lexical frequency when pitted against the non-scalar constraint MParse. In McClelland & Vander Wyk's (2006) study of interacting phonotactic constraints in English, a term β, analogous in effect to Albright's MParse, also improves model fit. Earlier, Frisch *et al*. (1997) used a constant term K pitted against a scalar constraint in order to derive logistic curves predicting the frequency of consonant pairs in Arabic roots from their similarity. Zuraw (2012), studying morphophonemic nasal mutation in Tagalog, finds that the Tagalog prefixes differ in their propensity to trigger nasal mutation, and stem-initial consonants differ in their propensity to undergo it; each prefix gives rise to its own sigmoid curve. Much earlier, Kroch (1989) presented a model in which a syntactic constraint rises in weight over time, overtaking a set of opposed statically weighted constraints, and thus creating a family of diachronic sigmoids; the model matches well with Kroch's diachronic data. For further discussion see Zuraw (2003).

7 Two non-restrictive approaches

We have thus far considered two models that work well in accounting for the Tommo So data (maxent, Noisy Harmonic Grammar), and also some models that work poorly (inverted exponentials and the stripped-down constraint sets of note 28). Here we critique two models that fit the Tommo So data perfectly, but at great cost in restrictiveness.

7.1 Morpheme indexation

It is likely that there are phonological processes triggered by specific morphemes or undergone by specific morphemes. Pater (2000, 2009b) suggests that particular markedness or faithfulness constraints can be indexed to particular morphemes. Work supporting this view includes Flack (2007), Gouskova (2007), Steriade (2008), Jurgec (2010) and Linzen *et al*. (2013).

We have found that if we affiliate a separate Agree constraint with each affix of Tommo So (along with one for roots), we can achieve a very good – indeed, perfect – fit to the data. For instance, we factored our system of Agree constraints to include 21 constraints in total – one for each combination of morphological type and harmony process; e.g. Agree $[\text{back}]_{\text{factitive}}$. Coupled with our three Ident constraints in (32) and implemented in maxent, this achieves an essentially perfect data fit (mean error 0.00000064).

As a reviewer points out, this 'would be a less restrictive model … since there would be no predicted connection between linear order of morphemes and frequency of alternation'. Indeed, the model is the least restrictive conceivable. No matter what data we gave it, constraint weights could be found that achieve a perfect fit.

A specific consequence is that morpheme indexation does nothing to explicate the essential Kiparskian generalisation that application frequency diminishes in outer morphological layers. We think our approach is a better theory, because it actually predicts this pattern. It remains a problem to explain why morpheme-specific markedness or faithfulness constraints do not commonly subvert the Kiparskian pattern; perhaps these constraints represent something of a 'last-resort' option of language learners, invoked only when more natural constraints cannot account for the data.

7.2 Domain-indexed constraints following a stringency hierarchy

A similar approach indexes constraints to morphological domains rather than individual morphemes. Kiparsky (1994), reanalysing the English *-t,d*-deletion pattern in §5, uses a model of this sort. The key idea is that 'a violation at any given level is necessarily a violation at all superordinate levels (but not conversely).' Applied to /t/-deletion, the pattern would look like (36). Such a pattern is a STRINGENCY HIERARCHY (Prince 1997, de Lacy 2004) and may be related to the Strong Domain Hypothesis of Lexical Phonology (Kiparsky 1984, Myers 1991; rules may 'turn off' at later levels but not 'turn on').

(36) *Stringency hierarchy of markedness on levels*

candidate	$*C{+}t_{root}$	$*C{+}t_{stem}$	$*C{+}t_{word}$
[cost]	*	*	*
[cos]t			
[los+t]		*	*
[los]t			
[toss#t]			*
[toss]#t			

Kiparsky stipulated that the ranking probability between each indexed markedness constraint and the opposing faithfulness constraints must be exactly the same. With this condition, the theory makes exactly the same predictions as the rule-based approach of Guy that inspired it. In particular, it works well for the English *-t,d*-deletion data – but by the same token, it fails badly when applied to Tommo So.

On the other hand, as a reviewer suggested, the approach might be adapted to Tommo So by abandoning the assumption that the ranking probability with respect to faithfulness must remain the same for each constraint in the stringency hierarchy. Pursuing this suggestion, we set up a hierarchy for the Tommo So AGREE constraints, illustrated here for Backness Harmony; forms are taken from §4.

(37) *Stringency hierarchy of markedness constraints for Backness Harmony*

candidate	locus of violation	Agree Lev 1 (root)	Agree Lev 2 (fact)	Agree Lev 3 (rev)	Agree Lev 4 (trans)	Agree Lev 5 (med)	Agree Lev 6 (caus)	Agree Lev 7 (perf)
kɔ́jɛ́	root	*	*	*	*	*	*	*
dʑɔ̀bɔ́-ndɛ́	factitive		*	*	*	*	*	*
úm-ílɛ́	reversive			*	*	*	*	*
ɔ́g-íɾɛ́	transitive				*	*	*	*
tɔ́m-íjɛ́	mediopass					*	*	*
kɛ́ɾɛ́-mɔ́	causative						*	*
bòd-ì	perfective							*

The domain-indexation model behaves similarly to the morpheme-indexation model just discussed: when coupled with a stochastic grammar framework, it can provide an exact match to the Tommo So data pattern. Moreover, unlike morpheme indexation, domain indexation at least does cover the central generalisation that application frequency descends going outward from the stem.[23] However, it also suffers from insufficient restrictiveness, as it can derive *any* non-ascending quantitative pattern.[24] For instance, it can derive patterns that descend in a straight line, or that asymptote at values other than one or zero (e.g. starting or ending with a level sequence at 0.5); for graphical demonstrations see the supplementary materials. In contrast, our scalar model is tightly constrained by the mathematical generalisations given in §6.6; it generates only symmetrical descending curves that asymptote at one and zero.

7.3 Summary and evaluation of all models

In this article, we have considered multiple models to account for the Tommo So vowel-harmony data. By far the most successful were those based on scalar constraints and Harmonic Grammar. Both the maxent and Noisy Harmonic Grammar versions achieved a close fit to the data. Poor fits, on the other hand, were obtained under a number of different approaches. The inverted-exponential model (§5) was one such case, whether or not an intercept term is included (note 22). Our own Harmonic Grammar models also performed poorly if we attempted to

---

[23] This is true under the standard assumption that negative weights are not allowed; i.e. we do not want constraints to reward violations.

[24] Specifically, we can always set the weights to match the data perfectly. The rate of harmony in Level 1 (perfective) is determined solely by the ranking or weighting of Ident *vs.* $\text{Agree}_{\text{Level 1}}$; this can be matched by choosing the appropriate ranking probability or weight. The rate of harmony in Level 2 (causative) will be (by hypothesis) the same as or higher than that in Level 1; this can be matched by selecting a suitable weight or ranking value for $\text{Agree}_{\text{Level 2}}$, creating an additional penalty on the non-harmonised candidate. Proceeding inductively through the hierarchy, we can obtain an exact match for every level.

simplify them by removing the possibility of language-specific weights for either faithfulness or markedness (note 22). Lastly, both the morpheme-indexation approach (§7.1) and domain indexation with stringency (§7.2) fit the data perfectly, but in a sense they are uninteresting, because they can fit either any data pattern whatsoever (morpheme indexation) or any descending pattern (domain indexation). Table V summarises, giving error values for each model.

		harmony errors			
		total	Low	Back	ATR
core model in maxent	§6.5	0.012	0.012	0.022	0.000
core model in NHG	§6.7	0.009	0.005	0.021	0.000
stochastic OT with exploded constraints	note 21	0.009	0.005	0.021	0.000
classic inverted exponentials	§5	0.181	0.249	0.134	0.157
inverted exponentials plus intercept	note 22	0.172	0.223	0.134	0.156
core model, fixed markedness	note 22	0.076	0.068	0.042	0.107
core model, fixed faithfulness	note 22	0.090	0.123	0.055	0.080
morpheme indexation	§7.1	no errors, but would fit any data pattern			
domain indexation	§7.2	no errors, but would fit any non-ascending data pattern			

Table V
Summary of models of the Tommo So vowel-harmony data.

8 Further issues

8.1 What kind of variation?

Variation in phonology takes several forms. For instance, it can be either interspeaker or intraspeaker variation. In the former case, every speaker always produces the same outputs, and the appearance of variation in the data as a whole is merely the result of mixing in data from different speakers. We think the variation we describe here is intraspeaker variation; there is no independent evidence of dialect differences among the language consultants, and the collected data include numerous instances in which the same speaker said the same word on different occasions with different harmony; see e.g. (9), (10) and (19).

Intraspeaker variation involves either types or tokens. In type variation, words or stems of similar phonological make-up behave differently on an idiosyncratic, lexically determined basis, but each one is always pronounced the same. In token variation, all words or roots that are phonologically eligible for variation vary from one speaking occasion to the next. In light of our observations, Tommo So cannot be a case of pure type variation, but it is possible that *some* words are lexically listed (or diacritically marked) as always having harmony or non-harmony. Such forms are difficult to detect in our data, and the results of the following statistical inquiry are inconclusive.

Test 1. One way of testing for the presence of listed forms is to check whether there are more all-harmony and no-harmony words among the population of varying forms than could arise by chance. We tested this with a Monte Carlo simulation (Mooney 1997). In our corpus there are 155 types that have a combination of root + suffix with a harmony rate above zero and below one, and with more than one attested token; these are the types in which token variation can, in principle, be detected. The total number of tokens of these types is 723. To execute one single Monte Carlo run on one single token, we toss a simulated biased coin whose probability of heads is equal to the theoretical probability of harmony, taken from Table III. Recording this outcome, we repeat the process for all 723 tokens. Examining the batch of pseudo-data thus created, we count the number of types whose tokens come out all-harmony, and the number of types that come out no-harmony. The whole process is repeated 100,000 times, yielding an approximate probability distribution for the expected number of all-harmony and no-harmony types. Lastly, we examine where the real counts (66 no-harmony, 48 all-harmony) fall in this distribution. The total fraction of Monte Carlo trials that have 66 or more no-harmony forms, or 48 or more all-harmony forms, divided by 100,000, yields a p-value, telling us the probability that these numbers could arise by chance. The results are highly significant: $p < 0.00001$ in both cases. In other words, our test shows that there are far more all-harmony and no-harmony cases then could ever arise by chance.

In principle, this could be taken as evidence that there is substantial type variation, in the form of root + suffix combinations that are memorised in some way. However, the picture is not so clear. Our experience in elicitation suggests that consultants are vulnerable to SELF-PRIMING: if they give one possible free-variation outcome then the odds are good that they will give the same one when asked a few minutes later. Under the self-priming effect, we might expect a statistically significant effect in the test just given, even if Tommo So words are not listed with their harmony outcomes.

Moreover, if there *are* type effects in our data, it is strikingly hard to locate the individual words that are responsible. This is what we found in Test 2.

Test 2. The goal is to find the words that display a great deal or very little harmony in the context of the overall probability predicted for them by our analysis. To do this, we consult the cumulative binomial distribution for (at least/fewer than) *m* harmony outcomes when there are *n* tokens (Mosteller *et al*. 1970: 138–145). In fact, very few words pass this test, even with a liberal 5% significance criterion.[25] One such example is /dùló-íjɛ́/ (turn around-MEDIOPASS) ‘return’, which appears eight times in the corpus as [dùlíjó] and never as [dùlíjé]. The probability of Backness Harmony in the mediopassive is 0.44; hence the probability of unanimity for [dùlíjó] is 0.44^8, or 0.0014; this is unlikely to be accidental.

Curiously, all nine words that pass this test at the 0.01 significance level (such as [dùlíjó]) are words with invariant harmony. To explain this, we suggest that these words have been restructured and are treated synchronically as roots (which in Tommo So virtually always undergo harmony; §4.1). A possible English analogue (Kiparsky 1977: 222) is the exceptional compound *high school* [ˈhʌɪ skul], treated as a single stem and thus eligible for the raising process /aɪ/ → [ʌɪ] / __ [−voice]; the normal outcome is seen in *pie school* [ˈpaɪ skul].

Obviously, more data would be needed to make further progress on the type/token question. For now, we think the best conjecture is that type variation may play a quite small role: we cannot prove that the overall disparities (Test 1) are not due merely to self-priming, and the very few cases where the type variation is clearly localisable to particular words (Test 2) share the all-harmony property, suggesting that the only mechanism at play is the treatment of a small number of historically polymorphemic words as monomorphemic.

8.2 Opacity with Hiatus Resolution

ATR Harmony in Tommo So is opaque: it is in a counterbleeding relationship with Hiatus Resolution. In roots with an initial high vowel and a final mid vowel, the only possible ATR trigger is the final vowel, since only mid vowels may trigger ATR Harmony (§4.4). When a vowel-initial suffix is added to such a root, this final vowel deletes, yet the suffix vowel agrees in [ATR] as though it were present, as shown in (38).

(38) a. *Deleted* [+ATR] *root vowel: suffix is* [+ATR]
/dìdé-ílɛ́/ (prop.up-REVERS) → [dìdílé] ‘remove prop’
/kúmbó-íɾɛ́/ (fist-TRANS) → [kúmbíɾó] ‘put in fist’
/wìgílé-íjɛ́/ (swing-MEDIOPASS) → [wìgílíjé] ‘swing’

b. *Deleted* [−ATR] *root vowel: suffix is* [−ATR]
/dìŋɛ́-ílɛ́/ (tie-REVERS) → [dìŋílɛ́] ‘untie’
/túŋɔ́-íɾɛ́/ (kneel-TRANS) → [túŋíɾɔ́] ‘make kneel’
/tímbɛ́-íjɛ́/ (stack-MEDIOPASS) → [tímbíjɛ́] ‘be stacked’

As these examples show, the suffix’s [ATR] value is determined by a vowel no longer present on the surface. In a rule-based analysis, this result would be obtained by ordering ATR Harmony before Hiatus Resolution.

In constraint-based phonology, there are several theories that would permit the derivation of these opaque cases. Employing autosegmental

---

[25] It is in fact extremely liberal because we are ‘fishing’; looking for the same outcome over and over again without adjusting the significance criterion.

theory, we could let vowel deletion strand a floating Root node (Clements 1985). Provided that this node is visible to the AGREE constraints, opacity would result. Opacity could also be obtained by framing our analysis within various standard approaches within OT: Sympathy Theory (McCarthy 1999), Turbidity Theory (Goldrick 2000), Stratal OT (Bermúdez-Otero 1999, Kiparsky 2000, 2003), Candidate Chain Theory (McCarthy 2007). Since the analysis of opacity is tangential to our main concerns, we will not explore the matter further.

9 Conclusion

We list what we take to be the main results of this work.

First, Tommo So vowel harmony is a clear case of the Kiparskian generalisation that phonological processes 'turn off' as they extend into the outer reaches of morphology. With a rich set of ordered affixes, Tommo So forms a test case for formal theories that seek to account for this phenomenon.

Second, Guy's inverted-exponential theory, though simple and principled, does not provide a good fit to the Tommo So data. Other models of phonology–morphology interactions, namely morpheme-indexed constraints and domain-indexed constraints with a stringency hierarchy, are able to provide an essentially perfect match, but this result is uninformative because these models are unrestrictive.

Third, Harmonic Grammars (both maxent and Noisy Harmonic Grammar), augmented to include scalar constraints, provide a good fit to our data. Moreover, these theories are restrictive; they inherently generate sigmoid probability functions and can only be used to model curves that fit this family.

This said, we judge that our work remains speculative until a greater number of 'petering out' systems have been located and documented with enough data to do quantitative analysis. We hope the method described here will facilitate the task of taking on new cases.

REFERENCES

Albright, Adam (2012). Additive markedness interactions in phonology. Ms, MIT. Available (January 2016) at http://www.linguistics.ucla.edu/colloquia/papers/AlbrightAdditiveMarkedness2012.pdf.

Baković, Eric (2000). *Harmony, dominance, and control*. PhD dissertation, Rutgers University.

Bermúdez-Otero, Ricardo (1999). *Constraint interaction in language change: quantity in English and Germanic*. PhD dissertation, University of Manchester & University of Santiago de Compostela.

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* **32**. 45–86.

Boersma, Paul & Joe Pater (2016). Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. In McCarthy & Pater (2016). 389–434.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook* **2**. 225–252.

Clements, G. N. & Engin Sezer (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*. Part 2. Dordrecht: Foris. 213–255.

Cole, Jennifer & Charles Kisseberth (1994). An optimal domains theory of harmony. *Studies in the Linguistic Sciences* **24**. 101–114.

de Lacy, Paul (2004). Markedness conflation in Optimality Theory. *Phonology* **21**. 145–199.

Della Pietra, Stephen, Vincent Della Pietra & John Lafferty (1997). Inducing features of random fields. *IEEE Transactions: Pattern Analysis and Machine Intelligence* **19**. 380–393.

Fasold, Ralph W. (1972). *Tense marking in Black English: a linguistic and social analysis*. Arlington: Center for Applied Linguistics.

Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *LI* **38**. 749–758.

Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* **18**. 7–44.

Frisch, Stefan A., Michael B. Broe & Janet B. Pierrehumbert (1997). Similarity and phonotactics in Arabic. Available as ROA-223 from the Rutgers Optimality Archive.

Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe (2004). Similarity avoidance and the OCP. *NLLT* **22**. 179–228.

Fruehwald, Josef T. (2012). Redevelopment of a morphological class. *University of Pennsylvania Working Papers in Linguistics* **18:1**. 77–86. Available at http://repository.upenn.edu/pwpl/vol18/iss1/10.

Goldrick, Matthew (2000). Turbid output representations and the unity of opacity. *NELS* **30**. 231–245.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.

Gouskova, Maria (2007). The reduplicative template in Tonkawa. *Phonology* **24**. 367–396.

Guy, Gregory R. (1980). Variation in the group and the individual: the case of final stop deletion. In Labov (1980). 1–36.

Guy, Gregory R. (1991). Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* **3**. 1–22.

Hantgan, Abbie & Stuart Davis (2012). Bondu-so vowel harmony: a descriptive analysis with theoretical implications. *Studies in African Linguistics* **41**. 187–212.

Hargus, Sharon (1988). *The lexical phonology of Sekani*. New York: Garland.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**. 379–440.

Hochstetler, J. Lee, J. A. Durieux & E. I. K. Durieux-Boon (2004). *Sociolinguistic survey of the Dogon language area*. SIL International. Available (January 2016) at http://www-01.sil.org/silesr/2004/silesr2004-004.pdf.

Inkelas, Sharon (1993). Nimboran position class morphology. *NLLT* **11**. 559–624.

Jesney, Karen (2011). Licensing in multiple contexts: an argument for Harmonic Grammar. *CLS* **45:1**. 287–301.

Jesney, Karen & Anne-Michelle Tessier (2011). Biases in Harmonic Grammar: the road to restrictive learning. *NLLT* **29**. 251–290.

Jurgec, Peter (2010). Disjunctive lexical stratification. *LI* **41**. 149–161.

Katamba, Francis (2004). *Morphology: critical concepts in linguistics*. London & New York: Routledge.

Kaun, Abigail R. (1995). *The typology of rounding harmony: an optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.

Kimper, Wendell (2011). *Competing triggers: transparency and opacity in vowel harmony*. PhD dissertation, University of Massachusetts Amherst.

Kiparsky, Paul (1973). Phonological representations. In Osamu Fujimura (ed.) *Three dimensions of linguistic theory*. Tokyo: TEC. 3–135.

Kiparsky, Paul (1977). The rhythmic structure of English verse. *LI* **8**. 189–247.

Kiparsky, Paul (1982). Lexical phonology and morphology. In The Linguistic Society of Korea (ed.) *Linguistics in the morning calm*. Seoul: Hanshin. 3–91.

Kiparsky, Paul (1984). On the lexical phonology of Icelandic. In Claes-Christian Elert, Iréne Johansson & Eva Strangert (eds.) *Nordic Prosody III*. Stockholm: Almqvist & Wiksell. 135–164.

Kiparsky, Paul (1994). An OT perspective on phonological variation. Handout from Rutgers Optimality Workshop 1993 and New Ways of Analyzing Variation 1994, Stanford University. Available (January 2016) at http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf.

Kiparsky, Paul (2000). Opacity and cyclicity. *The Linguistic Review* **17**. 351–365.

Kiparsky, Paul (2003). Syllables and moras in Arabic. In Caroline Féry & Ruben van de Vijver (eds.) *The syllable in Optimality Theory*. Cambridge: Cambridge University Press. 147–182.

Kirchner, Robert (1993). Turkish vowel disharmony in Optimality Theory. Paper presented at the Rutgers Optimality Workshop, Rutgers University.

Kroch, Anthony S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* **1**. 199–244.

Labov, William (ed.) (1980). *Locating language in time and space*. New York: Academic Press.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 388–395.

Linzen, Tal, Sofya Kasyanenko & Maria Gouskova (2013). Lexical and phonological variation in Russian prepositions. *Phonology* **30**. 453–515.

Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *NLLT* **17**. 267–302.

McCarthy, John J. (1999). Sympathy and phonological opacity. *Phonology* **16**. 331–399.

McCarthy, John J. (2004). Headed spans and autosegmental spreading. Ms, University of Massachusetts, Amherst. Available as ROA-685 from the Rutgers Optimality Archive.

McCarthy, John J. (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Sheffield & Bristol, Conn.: Equinox.

McCarthy, John J. & Joe Pater (eds.) (2016). *Harmonic Grammar and Harmonic Serialism*. London: Equinox.

McCarthy, John J. & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249–384.

McClelland, James & Brent C. Vander Wyk (2006). Graded constraints on English word forms. Ms, Carnegie Mellon University. Available (January 2016) at http://psych.stanford.edu/~jlm/papers/GCEWFs_2_18_06.pdf.

McPherson, Laura (2013). *A grammar of Tommo So*. Berlin & Boston: De Gruyter Mouton.

Mooney, Christopher Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Mosteller, Frederick, Robert E. K. Rourke & George B. Thomas (1970). *Probability with statistical applications*. 2nd edn. Reading, Mass.: Addison-Wesley.

Myers, Scott (1991). Structure Preservation and the Strong Domain Hypothesis. *LI* **22**. 379–385.

Nesbitt, C. F. (1984). *The linguistic constraints on a variable process: /t,d/ deletion in Sydney speech*. BA thesis, University of Sydney.

Neu, Helene (1980). Ranking on constraints on /t,d/ deletion in American English: a statistical analysis. In Labov (1980). 37–54.

Nida, Eugene (1949). *Morphology: the descriptive analysis of words*. 2nd edn. Ann Arbor: University of Michigan Press.

Pater, Joe (2000). Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* **17**. 237–274.

Pater, Joe (2008). Gradual learning and convergence. *LI* **39**. 334–345.

Pater, Joe (2009a). Weighted constraints in generative linguistics. *Cognitive Science* **33**. 999–1035.

Pater, Joe (2009b). Morpheme-specific phonology: constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological argumentation: essays on evidence and motivation*. London: Equinox. 123–154.

Pater, Joe (2016). Universal Grammar with weighted constraints. In McCarthy & Pater (2016). 1–46.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* **27**. 77–117.

Prince, Alan (1997). Paninian hierarchies. Ms, Rutgers University.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.

Rhodes, Russell (2012). Vowel harmony as Agreement by Correspondence. *Annual Report of the UC Berkeley Phonology Laboratory*. 138–168.

Ringen, Catherine O. & Orvokki Heinämäki (1999). Variation in Finnish vowel harmony: an OT account. *NLLT* **17**. 303–337.

Ringen, Catherine O. & Robert M. Vago (1998). Hungarian vowel harmony in Optimality Theory. *Phonology* **15**. 393–416.

Rose, Sharon & Rachel Walker (2004). A typology of consonant agreement as correspondence. *Lg* **80**. 475–531.

Ryan, Kevin M. (2010). Variable affix order: grammar and learning. *Lg* **86**. 758–791.

Santa Ana, Otto (1991). *Phonetic simplification processes in the English of the barrio: a cross-generational sociolinguistic study of the Chicanos of Los Angeles*. PhD dissertation, University of Pennsylvania.

Smolensky, Paul (1986). Information processing in dynamical systems: foundations of Harmony Theory. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (eds.) *Parallel Distributed Processing: explorations in the micro-structure of cognition*. Vol. 1: *Foundations*. Cambridge, Mass.: MIT Press. 194–281.

Smolensky, Paul & Géraldine Legendre (eds.) (2006). *The harmonic mind: from neural computation to optimality-theoretic grammar*. 2 vols. Cambridge, Mass.: MIT Press.

Steriade, Donca (2008). A pseudo-cyclic effect in Romanian morphophonology. In Asaf Bachrach & Andrew Nevins (eds.) *Inflectional identity*. Oxford: Oxford University Press. 313–359.

Sussman, Harvey M., Kathryn A. Hoemeke & Farhan S. Ahmed (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *JASA* **94**. 1256–1268.

Tagliamonte, Sali & Rosalind Temple (2005). New perspectives on an ol' variable: (t,d) deletion in British English. *Language Variation and Change* **17**. 281–302.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Walker, Rachel (1998). *Nasalization, neutral segments and opacity effects*. PhD dissertation, University of California, Santa Cruz.

Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* **30**. 945–982.

Wilson, Colin & Benjamin George (2009). Maxent grammar tool. Software package. Available at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool.

Zuraw, Kie (2003). Probability in language change. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.) *Probabilistic linguistics*. Cambridge, Mass.: MIT Press. 139–176.

Zuraw, Kie (2012). Quantitative patterns of constraint interaction. Paper presented at the 30th West Coast Conference on Formal Linguistics, University of California, Santa Cruz.