12

Phonological Regularity and Breakdown. An Account of Vowel Length Leveling in Middle English

Charles Yang

In Middle English, vowels in stressed syllables had length alternation under certain conditions that has been lost in Modern English. For example, the singular–plural pair *whal–whāles* 'whale–whales' had a short–long alternation and *crādel–cradeles* 'cradle–cradles' had a long–short alternation in Middle English, but there is no difference in their vowel length in Modern English. Building on traditional analyses (e.g., Prokosch, 1939), Lahiri and Dresher (1999) argue that the vowel length alternation in Middle English predictably follows from the interactions of several phonological processes, some shared across the West Germanic family and others unique to English. However, independent phonological changes

It is with pleasure and gratitude to dedicate the present paper to Don Ringe for his tutelage, support, and friendship. This work stems from some loose ideas about a decade ago: I would like to thank Aaron Ecay for his help with the historical corpus search at the time. Carola Trips and Jordan Kodner and the editors of the present volume also provided helpful comments and suggestions on an early draft.

C. Yang (✉)
University of Pennsylvania, Philadelphia, PA, USA
e-mail: charles.yang@ling.upenn.edu

J. F. Eska et al. (eds.), *The Method Works*,
https://doi.org/10.1007/978-3-031-48959-4_12

in the late Middle English period obscured the phonological condition on vowel length, which became sufficiently obscure for the learner to uncover. Subsequently, the vowel length was leveled: there is no alternation in Modern English except for a few archaic remnants such as *staff–staves*. Curiously, words that belong to the same historical classes leveled differently. For example, in Middle English, *path–pāthes* 'path–paths' patterned like *whal–whāles*, and *sādel–sadeles* 'saddle–saddles' patterned like *crādel–cradeles*. In Modern English, however, *path* has a short vowel and *whale* has a long vowel, whereas *cradle* has a long vowel and *saddle* has a short vowel.

In this note, I follow Lahiri and Dresher's approach, but supplement it with a learning model known as the Tolerance Principle (TP; Yang, 2005, 2016). The TP is a mathematical measure that formalizes exactly what it means for a phonological process to be systematic and predictable, and when it collapses. I show that the quantitative composition of the learning data in late Middle English, indeed, could not support vowel length alternation across the board, thereby providing a more precise causal explanation for its demise. I also make some suggestions on the directionality of leveling which may appear quite arbitrary at first glance as in the example of *path–whale* and *cradle–saddle* examples above. But first, let us review some of the essential facts and Lahiri and Dresher's proposal.

1 The Leveling of Vowel Length in Middle English

The Open Syllable Lengthening (OSL) rule, which lengthens vowels in stressed open syllables, was a general property of West Germanic languages including Middle English (Prokosch, 1939). From the late Old English period through at least some stages of Middle English, the language also developed the Trisyllabic Shortening (TSS) rule (Lass, 1992; Wright & Wright, 1928). TSS shortens the longer vowel if it is followed by two or more syllables, at least one of which was unstressed. It is productively applied to all vowels in Middle English and retains some productivity even in Modern English (Chomsky & Halle, 1968). The relevant patterns in Table 12.1 are taken from Lahiri and Dresher's paper, which also presents arguments that TSS followed the application of OSL.

Table 12.1 Expected effects of Open Syllable Lengthening (OSL) and Trisyllabic Shorting (TSS) in Middle English (from Lahiri & Dresher, 1999, p. 680)

	SG	PL	SG	PL	SG	PL	SG	PL
Old English	hǣring	hǣringas	hamor	hamoras	stȳpel	stȳpelas	beofor	beoferas
OSL	—	—	hāmor	hāmores	—	—	bēver	bēveres
TSS	—	hæringes	—	hamores	—	stypeles	—	beveres
Expected	hǣring	hǣrings	hāmor	hamors	stȳpel	stypels	bēver	bevers
Modern English	herring	herrings	hammer	hammers	steeple	steeple	beaver	beavers

These rules facilitate a systematic mapping from the underlying representation of the words to their surface inflections. Consider *hāring* 'herring', which has an underlying long vowel, and *hamor* 'hammer', which has an underlying short vowel. Their surface realizations, however, have the same vowel length, thanks to the vowel length alternation regulated by OSL and TSS: long in the singular and short in the plural. The alternation was lost in later periods of English. In Modern English, almost all nouns have the same vowel (length) in both singular and plural, and their underlying representations—not a particularly useful notion now, unlike in Old English—are largely identical to the surface forms.

Clearly, some kind of leveling had taken place in the intervening centuries from Old English to Modern English. Note further that the leveling went both ways. In Table 12.1, words such as *herring* and *hammer* have preserved the vowel length (short) in the plural, whereas *steeple* and *beaver* have preserved the vowel length (long) in the singular. The aim of this chapter is to understand why and how leveling took place and to predict, when possible, which vowel length (in singular or plural) was retained.

To do so, we need to examine the major phonological classes of nouns and how they are affected by OSL and TSS. Consider the examples in Table 12.2 (Lahiri & Elan Dresher, 1999, p. 690).

A list of example words from these four classes can be found in Sects. 3.1 and 3.2, when I discuss what turns out to be quite different leveling processes for them. For the moment, let us review the basic facts about the way they changed.

The class with vocalic endings in both singular and plural (Table 12.2a) are the neuter nouns that make up about 25% of the Old English vocabulary (Hogg, 1992, p. 126). These were subject to OSL, but not TSS, due

Table 12.2 Expected effects of OSL and TSS in Middle English on the vowel length of Old English nouns

	Old English paradigm				Expected surface length	
	Stem vowel length	Endings	SG	PL	OSL/TSS	Gloss
(a)	short	V-V	talu	tala	L-L	'tale'
(b)	short	∅-V	hwæl	hwalas	S-L	'whale'
(c)	short	∅-V	beofor	beoferas	L-S	'beaver'
(d)	long	∅-V	hǣring	hǣringas	L-S	'herring'

to their bisyllabic nature, and, therefore, surface as L-L (long-long) in both singular and plural. By Middle English, the plural suffix *-es* had already become dominant irrespective of the historical classes (Lass, 1992, p. 109) so *talu–tala*, therefore, would become *tāle–tāles* with a pronounced schwa. These nouns generally retained the long vowel, including those from the major *a*-, *e*-, and *o*-stem classes such as *nāma* 'name', *bēdu* 'bead', and *clōca* 'cloak'. An important characteristic of these nouns is that their singular form had a vocalic ending and, in particular, a schwa by the Middle English period as in *tale*, *name*, *bede*, and *cloke*.[1] This characteristic will prove critical to the relative orderliness of their vowel length change as I discuss in Sect. 3.1.

By contrast, the other three classes in Table 12.2 leveled to both long and short vowels in a seemingly chaotic fashion, the focus of my discussion in Sect. 3.2. The old monosyllabic *a*-stem nouns (Table 12.2b) have an underlying short vowel. They have a closed syllable in the singular and an open syllable in the plural, resulting in an S–L (short–long) alternation due to OSL. Again, TSS does not apply. These nouns lost the length alternation with some leveling to short, e.g., *back*, while others to long, e.g., *blade*. The disyllabic nouns with an underlying short vowel (Table 12.2c) are bisyllabic in the singular, but trisyllabic in the plural, due to the addition of the inflectional suffix. These are, thus, subject to OSL in the singular and both OSL and TSS in the plural, resulting in an L–S alternation. Some of these nouns leveled to short, e.g., *botm* 'bottom', while others leveled to long, e.g., *æcer* 'acre'. Similarly, the disyllabic

[1] Their orthography in Modern English follows the so-called silent-e rule: word-final *e* is silent, which reflects the loss of schwa by the mid-fifteenth century, and the preceding vowel is long.

nouns with an underlying long vowel (Table 12.2d) are subject to TSS in the plural, resulting in L–S alternation. Again, leveling to both lengths can be observed: *bēacon* 'beacon' to long and *bōsm* 'bosom' to short.

What led to the breakdown of vowel length alternation and subsequent leveling? According to Lahiri & Dresher (1999), the state of confusion arose due to the loss of the inflectional vowel in late Middle English (e.g., Lass, 1992; Minkova, 1982). The schwa in plurals was dropped after vowel-final stems, as well as in polysyllabic words, eventually leading to the loss of schwa generally (except following sibilants): the suffix became just *-s*, which was assimilated for voicing to the final segment. Importantly, the loss of the inflectional vowel resulted in the loss of a syllable so that TSS would not be applicable to words such as those in Tables 12.2c and d. As I illustrate with some examples in Sect. 3, this independent change of schwa loss made it impossible to make consistent inference about vowel length. Lahiri and Dresher (1999) suggest that the unpredictability of length alternation led the learner to postulate a uniform length of the vowels in the singular and plural: 'On our account, language learners despair of a rule, and opt instead to choose a consistent vowel quantity on a word-by-word basis' (698).

In what follows, I review the Tolerance Principle (TP; Yang, 2005, 2016), a mathematical principle that governs how language learner detects regularities in linguistic data (Sect. 2). I then apply the TP to the present case to provide a quantitative argument as to why leveling took place. I agree with Lahiri and Dresher (1999): leveling resulted from the unpredictability of vowel length following the loss of the inflectional schwa. The TP provides a concrete measure of what exactly it means to be unpredictable, a notion only alluded to in their analysis. The nouns in one of the phonological classes did in fact behave predictably (Sect. 3.1). For the rest, which genuinely had no regularities for the learner to detect, there still seems to be an interesting pattern in their directionality of leveling (Sect. 3.2), which on first look appears quite arbitrary, with some nouns retaining the length in the singular and others retaining the length in the plural.

2 Quantifying Regularity: The Tolerance Principle

As young children vividly illustrate in the classic Wug test (Berko, 1958), the ability to extend rules to novel items is a key feature of language.[2] But rule formation in language acquisition takes time. Children need to acquire a vocabulary from which rules can be established: word learning takes place very slowly in the early stages, peaking at just over 1000 at age three (Bornstein et al., 2004; Fenson et al., 1994). Additionally, since rules are almost always laden with exceptions, they can only become productive when the exceptions are overcome.

It has long been recognized, and from a broad range of theoretical perspectives (e.g., Aronoff, 1976; Bybee, 1995; Nida, 1949; Plunkett & Marchman, 1993), that the productivity of rules must overcome the exceptions. The Tolerance Principle is a precise theory of how much supporting evidence is necessary for a rule to become productive and generalized.

(1) The Tolerance Principle (TP):
Let a rule R be defined over a set of N items. R generalizes if and only if e, the number of items not supporting R, does not exceed θ_N:

$$e \leq \theta_N = \frac{N}{\ln N}$$

If e exceeds θ_N, then the learner will simply memorize the input that follows R and does not generalize beyond: i.e., R is unproductive. In that case, when the speaker encounters a new item for which R is in principle eligible, they would be at a loss, resulting in gaps and ineffability (Björnsdóttir, 2023; Gorman & Yang, 2019; Halle, 1973; Yang, 2017).

Because of its simplicity, the TP has been effectively applied to a wide range of learning problems in phonology, morphology, and syntax. Recent work includes gender assignment and inflection in Icelandic (Björnsdóttir, 2021), noun diminutives in Dutch (van Tuijl & Coopmans,

[2] It is convenient to think of 'rules' as classical rewrite processes in linguistics, but the term is used here in a pre-theoretic sense to denote any mapping or pattern defined over a set of items.

2021), argument structure mappings in English (Pearl & Sprouse, 2021), verbal inflection variation in Frisian (Merkuur, 2021), possessive suffix in Northern East Cree (Henke 2023), and others. One of the key properties of the TP is its recursive application. If a set of words fails to yield a productive generalization, i.e., if no single rule covers a sufficiently large number of words as defined by θ_N, the learner may subdivide the words into distinct sets and seek productive generalizations recursively within. Recursive application of the TP enables the learner to detect 'nested' regularities such as the German noun plural system, where the selection of the suffix is conditioned on the subsets of nouns defined by phonological properties as well as grammatical gender (Wiese, 1996; Yang, 2016). Furthermore, the experimental finding that infants follow the TP in implicit learning tasks (Emond & Shi, 2021; see also Schuler et al., 2016) suggests that the TP is likely a formal principle of learning and generalization, one that is not restricted to rule formation in language.

A theory of productivity in language acquisition has immediate consequences for language change. A new linguistic form regardless of its origin—contact, innovation, or innate biological capacity—must be transmitted by generations of child learners in order to take hold (Halle, 1962; Lightfoot, 1979; Paul, 1920). When a rule is acquired as productive, its openness may assimilate new eligible members. When a rule is acquired as unproductive, the only connection between it and the words that follow it would be experience, i.e., rote learning. Reduced exposure would result in words drifting to other (productive) rules, akin to overregularization errors in child language. Under this view, the traditional notion of analogical change (e.g., leveling and extension) can be understood as words responding to changes in the productivity of rules they fall under (Yang, 2016, pp. 139–170).

Since the calibration of productivity under the TP depends on only two values, N and e, it is possible to develop *predictive* accounts of language change. Or more precisely, since we are dealing with the past, accounts that aim to show that certain attested changes in history were, in fact, inevitable. Successful case studies can be found in the inflection of past participles in Latin (Kodner, 2023), the reorganization of the English metrical stress system due to the influx of Latinate vocabulary (Dresher & Lahiri, 2022), contact-induced phonological change in the

city of Philadelphia (Sneller et al., 2019), the rise and fall of English past tense inflection (Ringe & Yang, 2022), and the development of psych-experiencer verbs in English in contact with French (Trips & Rainsford, 2022), among others.

Quite critical to any TP-based analysis is to obtain accurate measures of *N* and *e* such that productivity calculation can be made. These values can be precisely manipulated in artificial language learning studies even at the individual level (e.g., Schuler, 2017) and can also be reliably obtained when the relevant vocabulary set is small and their linguistic history is well understood. For example, Ringe and Yang (2022) studied the productivity of a strong verb class that has/had no more than 20 items. Their dates of attestation and usage patterns can be found in the OED and the historical databases such as the Penn-Helsinki Parsed Corpus of Early Modern English (Kroch et al., 2004) and the Parsed Corpus of Early English Correspondence (Taylor et al., 2006), allowing for fine-grained TP calculations and predictions. But the general problem remains: it is hard enough to obtain realistic acquisition data for living languages, what is to do with dead languages with perhaps only a few hundred thousand words of surviving text?

It turns out that the psychological condition on child language acquisition greatly mitigates the problem of data poverty. Recall that children learn the core of their grammar very early: major properties such as inflections, case marking, word order, and transformations are all in place by age three, a stage where the vocabulary size can reach just north of 1000 words and often even lower for many children. Furthermore, these vocabulary items are all among the most frequent items in the language, so children have a realistic chance of acquiring them (Goodman et al., 2008): high-frequency items are, of course, also those that are more likely to be preserved in historical documents. In an important contribution, Kodner (2019) demonstrates the methodological soundness of using historical data as an approximation of child input data at the time. Specifically, when we restrict the words to the most frequent ones—e.g., the top 1000—in child-directed corpora, as well as adult language materials including historical corpora, we obviously obtain very different words, but the *rules* that these words support are very similar. Under the TP, the productivity of rules may be the same even if they are derived

from very different words, as long as the ratio of exceptions (i.e., e/N) falls on the same side of the TP threshold (i.e., $1/\ln N$).

With these methodological considerations in mind, let us proceed to the vowel length leveling problem from the perspective of the Tolerance Principle.

3 What Leveled, Why, and Where To?

The phenomenon of vowel length alternation and its leveling can be framed in terms of productivity. What Lahiri & Dresher (1999) refer to as 'predictable' vowel lengths is interpreted as native speakers having learned the phonological rules at the time, as described in Tables 12.1 and 12.2, and, therefore, did not have to memorize the vowel length in surface forms (singular/plural) by rote learning. This is formally equivalent to what we mean by a productive rule. For example, Modern English past tense is predictable because the rule (add *-ed*) is productive, and children learn so, presumably by following the TP, when the number of irregular verbs they know falls below the threshold. In this section, I will apply the TP to show that, following the loss of the inflectional vowel, the predictability of vowel length alternation was, indeed, undermined.

To carry out such an analysis requires concrete lexical statistics. To do so, I again turn to Lahiri & Dresher (1999), who provide a list of words of 186 nouns to adduce quantitative support for their analysis. These authors are quite explicit that their data are not a full description of the language at the time. Nevertheless, their decision to select only the most common words may just prove appropriate for our learning-theoretic approach to change. Of course, it is impossible to directly study the acquisition of historical languages. But there are reasons to believe that vowel length alternation would be among the first phonological properties that children learn. For example, Icelandic is a language that has a similar vowel-lengthening process in stressed open syllables (Árnason, 1998). Icelandic-learning children learn vowel length (Masdottir, 2008) and associated inflection (Thordardottir et al., 2002) very accurately by two and one-half—and they must have done so on a very modest, but high frequency, vocabulary that probably contains little more than 186

nouns. Unless the proportion of exceptions in Lahiri and Dresher's word list were very different from that in late Middle English child learner's vocabulary—the *only* condition that invalidates the application of the TP—we can be content with using their data as a surrogate.

3.1 What Leveled and Why

Let us now examine how the loss of the inflected vowel in the plural suffix could lead to the breakdown of vowel length alternation. Consider the representative examples below from Lahiri & Dresher (1999, p. 698):

(2) a. Before the loss of inflected vowel:

SG	PL	SG	PL	SG	PL	SG	PL
stōn	stōnes	god	gōdes	bōdi	bodies	bēver	beveres

b. After the loss of inflected vowel:

SG	PL	SG	PL	SG	PL	SG	PL
stōn	stōns	god	gōds	bōdi	bodis	bēver	bevers

(2a) reflects the grammar of Middle English speakers before the loss of inflectional vowel. As noted before, the vowel length alternation was predictable. (2b) would be the output of speakers after the loss of the inflectional vowel, which took place in late Middle English, a change that affected not only the plural suffix but also the past tense suffix (Lass, 1992). These forms would result from removing the schwa from (2a), which can be viewed as schwa deletion taking place after the application of OSL and TSS.

For the monosyllabic word *stōn* 'stone' in (2b), which has an underlying long vowel, the loss of the inflectional schwa affects nothing as the vowel surfaces as long for both singular and plural. But the other three examples in (2b) become problematic.

Take the monosyllabic noun *god*, which had an underlying short vowel. Before the loss of the inflectional schwa, the vowel is short in the singular and is lengthened in the plural as the *-es* suffix attracts /d/ as its onset, leaving the vowel open and, thus, eligible for OSL. Both surface forms are consistent with the underlying stem vowel being short. After the loss of the inflectional schwa, /ds/ becomes the coda of the vowel.

Being in a closed syllable, the vowel should not be eligible for OSL and, thus, should be short. However, the learner would hear *gōds* as having a long vowel, presumably produced by adults who had added schwa loss to their grammar. The learner would have to conclude that the underlying vowel is long. However, when the learner heard the singular *god* with a short vowel, with no relevant phonological rule at play, they would have to conclude that the underlying vowel is short. A clear incongruence.

The situation for the bisyllabic *bōdi* 'bodi' and *bēver* 'beaver' is similarly problematic. Before the loss of the inflectional schwa, their length alternation is predictable. The vowel is lengthened by OSL in the singular and shortened by TSS (after OSL) in the plural. Afterward, the plural becomes bisyllabic and is, thus, not subject to TSS, but should surface as long via OSL. But the input data in (2b) shows that the vowel in the plural is short. Again, an incongruence.

Taken together, the language learner would not be in a position to discover a systematic correspondence that regulates vowel length, even though this does not affect all words. It is worth noting that by this stage of late Middle English, other aspects of noun inflection (e.g., gender and case) had completely eroded away (Allen, 1999; Lass, 1992), leaving singular and plural marking the sole source for learning alternations.

The TP offers a quantitative measure of the conundrum the language learner would face when presented with data such as (2b). To take a trivial example, Modern English nouns clearly have no vowel alternation of any sort. This is so despite the fact that a small minority of nouns actually have different vowels in the singular and plural, a matter of historical residue: *child–children*, *tooth–teeth*, *index–indices*, *locus–loci*, *man–men*, etc. But these constitute only a very small proportion of nouns. Based on a standard word frequency norm (Brysbaert & New, 2009), only 15 such plurals appear more than once per million: *bases*, *children*, *criteria*, *data* (*datum*), *feet*, *geese*, *graffiti* (*graffito*), *men*, *media* (*medium*), *mice*, *opera* (*opus*), *phenomena*, *teeth*, *vertebrae*, and *women*. These are nowhere near enough to disrupt the generalization that Modern English has no vowel alternation.

Let us now consider the lexical statistics of vowel length alternation adapted from Lahiri and Dresher (1999, pp. 691–692) and summarized in Table 12.3.

Table 12.3 Four groups of nouns classified by their inflectional ending in Old English (OE), their expected length alternation, and their vowel length in Modern English (Modern)

OE ending		Example	Expected length	Total	Modern short	Modern long
V-V	a-stem	talu 'tale'	L-L	46	3 (6.5%)	43 (93.5%)
	e-stem	bedu 'bead'	L-L	27	9 (33.3%)	18 (66.7%)
	o-stem	nosu 'nose'	L-L	24	3 (12.5%)	21 (87.5%)
	Ø-V	hwæl 'whale'	S-L	36	19 (53%)	17 (47%)
	Ø-V	beofor 'beaver'	L-S	33	19 (58%)	14 (42%)
	Ø-V	hæring 'herring'	L-S	19	10 (53%)	9 (47%)

The situation is clearly very different from Modern English. There are 186 words: any generalization over them cannot have more than 186/ln186 or 35 exceptions. While the slight majority of them in the top half of Table 12.3 have the same vowel length in both singular and plural (L–L), this does not hold for the other three classes in the bottom half, 88 in all, far exceeding the TP threshold. Therefore, the learner confronted with the data represented in Table 12.3 could not reach any coherent conclusion about vowel length: it certainly is not Modern English, which has no vowel alternation.

But there is another route forward. Recall the recursive use of the TP on subdivided vocabulary sets when no 'global' productivity emerges out of the full set. The nouns with vocal endings in Old English, e.g., *talu–tala*, and with a schwa ending in the singular in Middle English, e.g., *tale–tales*, do have a consistent vowel length (L–L) in the singular and plural, as shown in the top half of Table 12.3. Examples of these words, which appear in the *a*-, *e*-, and *u*-stem classes, are given in (3):

(3) Examples of disyllable nouns with vocalic endings in Old English and with a schwa ending in the singular in Middle English (Lahiri and Dresher 1999, p. 690):

a. Stem vowel /a/: *apa* 'ape', *blæse* 'blaze', *bracu* 'brake', *nama* 'name', *snaca* 'snake', *spada* 'spade', *staca* 'stake', stalu 'stale', *talu* 'tale'
b. Stem vowel /e/: *bedu* 'bead', *peru* 'bear', *cwene* 'queen', *slege* 'slay', *smeoru* 'smear', *spere* 'spear', *stæpel stepe* 'step', *tere/teoru* 'tar'
c. Stem vowel /o/: *cloca* 'cloak', *fola* 'foal', *nosu* 'nose', *smoca* 'smoke', *stole* 'stole', *stofu/-a* 'stove', *sopa* 'sup', *þrotal-u* 'throat'

These words are sharply contrasted with the nouns in the bottom half of Table 12.3, which end with a consonant in the singular, as well as problematic words such as *bōdi* 'body' discussed in (2b), whose singular ends in a full vowel. Therefore, the learner should be able to carve out a formal class ('singular schwa ending') and conclude that, for this subset, the vowel should be consistently long. As illustrated in Table 12.3, this is, indeed, what happened. The vast majority of the 97 nouns in this class have long vowel in Modern English, with a small number (15) of exceptions. Lahiri & Dresher (1999) suggest that some of the exceptions are only apparent, with several showing long/short variation before settling on short. The TP offers a more reassuring answer. Even if all 15 exceptions are genuine and have always appeared in the input data with a short vowel, they still fall under the TP threshold as $\theta_{97} = 21$ and, thus, would not undermine the conclusion that the vowel in this class is long.

However, the rest of the 88 nouns in the bottom half of Table 12.3 remain a mess. These nouns have a formal characterization in opposition to those in the top half: their singular ending in Middle English is not a schwa. Even though a majority of these (33 + 19 = 52) have a long vowel in the singular which shortens in the plural, they fail to clear the TP threshold as $\theta_{88} = 19$. As Table 12.3 shows, there is no clear pattern in the direction of leveling, approximately half to long and half to short. This state of affairs accords with Lahiri & Dresher's (1999) perspective, and our reinterpretation, that the learner fails to detect any systematic correspondence in vowel length and must resort to rote memorization on a word-by-word basis. In their discussion, Lahiri and Dresher observe that when language learners make decisions on the assignment of new words into declensional classes, the nominative singular may be given prominence (Lahiri & Dresher 1983), a familiar notion from the theory of markedness (Greenberg, 1966; Jakobson, 1932, 1971). For the three classes of nouns here, the privileged status of the singular would predict more instances of leveling to short for the *whale* class and to long for the *beaver* and *herring* class. But Table 12.3 shows that that is not the case: there is no clear pattern favoring the length in the singular.

While the absence of systematicity does take the learner into the realm of uncertainty, there may still be some discernible patterns to uncover.

3.2 Directionality of Leveling

In general, the singular as an inflection is used far more frequently than the plural. In the one-million-word Brown Corpus (Kučera & Francis 1967), the total frequency of singular nouns outnumbers that of plural nouns by a ratio of 2.5:1. In the 15 million words of child-directed English extracted from the CHILDES database (MacWhinney, 2000), singulars hold an even larger advantage, outnumbering plurals more than five times. The statistical dominance of the singular over the plural can be seen in historical data, as well: the average ratio of singular versus plural frequency in the Penn-Helsinki Parsed Corpus of Middle English (PPCME2; Kroch & Taylor, 2000) is 1.68:1. These statistics do lend support for the privileged status of the singular in leveling and other processes of historical change. The learner hears the singular more often: in the absence of systematic correspondences, they would assume the singular form to be the base.

Except when they do not, as in the Middle English noun classes that do not have a schwa singular ending, which shows no preference for the singular (Table 12.3, bottom half). Previous research has identified similar cases where words leveled to the plural rather than the singular. For example, Tiersma (1982) aims to characterize the nature of the privileged status when the plural trumps the singular in several attested changes: nouns that typically come in pairs or groups (e.g., body parts such as *arm*), those that are more frequently used in the plural (e.g., shells on the beach, citing Berman, 1981), those that are otherwise favored by cultural conventions (e.g., *bacteria*), etc. These considerations are clearly relevant for the understanding of language use and change more generally, but they are of limited value in the present case. It is conceivable that *god* is more prominent than *gods* in a monotheistic culture, leading to the vowel length to level in the direction of the singular (short) over the plural (long). But it is difficult to see how *gates* (long) would have an advantage over *gate* (short), or how *acres* (long) would win over *acre* (short), etc.

I put forward the following conjecture. For leveling of the type considered here, where the paradigm becomes incoherent, the learner fails to find productive generalizations for length alternation and is, thus, compelled to select a length in one of the forms as the base. The directionality does not reflect the inherent privilege of certain inflectional class, but the frequency of usage—but with an important twist.

Let us examine the three problematic classes in turn. Consider first the monosyllabic *a*-stem nouns in (4).

(4) Old English monosyllable *a*-stems with short vowels (Lahiri & Dresher 1999, p. 691):

a. Short in NE: back, bath, black, brass, broth, chaff, glass, god, grass, lock, lot, path, sap, shot, staff, swath, thatch, vat, wer [wolf]
b. Long in NE: bead, blade, coal, crate, dale, day, door, fare, gate, grave, hole, hope 'recess', meet, sole 'mud', way, whale, yoke

Of the 36 nouns in (4), 19 are leveled to short (i.e., singular) and 17 are leveled to long (i.e., plural). In the Brown Corpus, the words that leveled to short/singular (4a) have an average singular frequency of 49.6 and an average plural frequency of 13.3. In comparison, by contrast, the words that leveled to long/plural (4b) have an average singular frequency of 138.8 and an average plural frequency is 44.1. In both cases, the singular is considerably more frequent than the plural: no surprise. But a striking pattern emerges. Those that leveled to the plural are just more frequent across the board: their *plural* forms are almost as frequent as the *singular* forms of the short vowel words (**44.1** vs. **49.6**), and they are over three times more frequent than the plural frequency of those that leveled to the short vowel (44.1 vs. 13.3).

These patterns are also observed in the Middle English corpus (Kroch & Taylor, 2000). The nouns that leveled to short/singular (4a) have an average singular frequency of 14.3 (out of 1.1 million words) and an average plural frequency of 3.4. Those that leveled to long/plural (4b) have an average singular frequency of 74.1 and an average plural frequency of 14.8. Again, as in the Brown Corpus, the *plural* frequency of those leveled to the plural length is comparable to the *singular* frequency of those leveled to the singular length (**14.8** vs. **14.3**). Once again, as is the case more generally, the singular is almost always more frequent than the plural. If leveling were to favor to more frequent, one would expect most nouns to level to the singular by preserving the short vowel, contrary to the facts in Table 12.3. Rather, the observed change would be accounted for if the target of leveling must meet some kind of frequency threshold: that only forms above certain frequency—be they singular or plural—are eligible to serve as the target of leveling.

A frequency threshold may be a somewhat alien proposal in the study of language change, but it is rooted in the study of language acquisition and more specifically lexical learning. Children learn words very slowly. On the one hand, the world in which words are embedded is messy and complex: finding the meanings of words can be very challenging (Gleitman & Trueswell, 2020). On the other, learning a word requires repeated exposure as some kind of rote memorization is always involved (Goodman et al., 2008). But lexical memory has a ceiling: if the exposure to a word is sufficiently high, then the learner will successfully acquire it. A case in point is the acquisition of irregular verbs. After the productive '-ed' rule is acquired, over-regularization errors will ensue: more frequent irregular verbs (e.g., *think–thinked* as opposed to *thought*) tend to have lower error rates than less frequent ones (e.g., *draw–drawed* instead of *drew*); see Marcus et al. (1992), Maratsos (2000), Yang (2002) for discussion. But eventually, everyone learns the irregular verbs correctly, despite the fact that *thought* is more frequent than *drew* at every stage of language acquisition. A similar notion is the finding that words occurring at least once per million are generally known to all high school graduates (Nagy & Anderson, 1984): more frequent than that does not make a word 'more' known. A frequency threshold boils down to this: if you know a word, you know a word. And if you know a word—by hearing it enough—you know the length of the vowel in it.

The proposal of a frequency threshold is also consistent with the two disyllabic noun classes.

(5) Old English disyllabic nouns with short open syllables (Lahiri & Dresher 1999, p. 691):

a. Long in NE: æcer 'acre', bydel 'beadle', beofor 'beaver', cradol 'cradle', efes 'eaves', efen 'even', hæfen 'haven', hæsel 'hazel', hlædel 'ladle', mapul- 'maple', nacod 'naked', hræfn 'raven', stapol 'staple', tapor 'taper'
b. Short in NE: botm 'bottom', camel 'camel', canon 'canon', copor 'copper', fæder 'father', fæþm 'fathom', feþer 'feather', fetel 'fettle', hamor 'hammer', heofon 'heaven', hofel 'hovel', lator 'latter', ofen 'oven', oter 'otter', sadol 'saddle', seofen 'seven', sc(e)ofl 'shovel', wæter 'water', weder 'weather'

(6) Old English disyllabic nouns with long vowels (Lahiri and Dresher 1999, p. 692):

a. Long in NE: bēacon 'beacon', bītel, bīetel 'beetle', æfenn 'even(ing)', hūsl 'housel, Eucharist', stȳpl 'steeple', tācn 'token', brīdels 'bridle', fēfor 'fever', hæþen 'heathen'
b. Short in NE: bōsm 'bosom', brōþor 'brother', dēofol 'devil', fōdor 'fodder', hæring 'herring', mōdor 'mother', rædels 'riddle', spātl 'spattle, saliva', þȳmel 'thimble', wæpen 'weapon'

As discussed earlier, both classes would show L–S alternation in Middle English, and the loss of the inflectional vowel in late Middle English led to the breakdown of predictable vowel length alternation. In the Brown Corpus, the nouns that leveled to long/singular (i.e., 5a and 6a) have an average singular frequency of 12.5 and an average plural frequency of 7.8.[3] The nouns that leveled to short/plural (i.e., 5b and 6b) have an average singular frequency of 46.6 and an average plural frequency of 12.9. Again, we see that the words that leveled to the plural length are on average much more frequent than those that leveled to the singular length. And again, the plural frequency of (5b) and (6b) is comparable to the singular frequency of (5a) and (6a): **12.9** versus **12.5**. Results from PPCME2 are similar. The nouns that leveled to the singular are quite infrequent: the singulars average 1.8 occurrences and the plurals 1.1. By contrast, those that leveled to the plural length have an average singular frequency of 16.7 and plural frequency of **2.7**, which is a bit higher than the *singular* frequency (**1.8**) of those that leveled to long, although the data is quite sparse.

For both the monosyllabic *a*-stem nouns (4) and the disyllabic nouns (5–6), those that leveled to the plural length are generally high frequency across the board. Even when their singular frequencies are still higher, their plural frequencies are high enough so the plural forms become eligible as target of leveling. The privileged status of the singular in leveling and change reflects the typical situation where the singular is more

[3] It is not surprising that bisyllabic words are less frequent than monosyllabic words reviewed earlier: longer words are less frequent than shorter ones (Caplan et al., 2020; Zipf, 1949).

frequent than the plural. But if the plural is already in the upper echelon of frequency, it can effectively neutralize the singular's advantage and become a target for leveling, even though the singular may be more frequent still.

4 Conclusion and Prospects

This has been a thought experiment. We imagined ourselves as child learners in the late period of Middle English. Armed with independently motivated mechanisms of language acquisition, we wondered what kind of grammar could have been acquired, one which would differ from our parents as the learning data had changed. The thought experiment is also enabled by the uniformitarian assumption. The psychological mechanism for language acquisition has not changed in the past few hundred years, nor has the ecological condition of language acquisition: children, then as now, learn their grammar from a fairly small set of highly frequent words. These considerations collectively allow us to build on previous scholarship (Lahiri & Dresher 1999) and develop precise hypotheses about change, which can then be verified on the available historical data. The methods are general and can be extended to other empirical studies. For example, we have assumed that if the child learner fails to detect any systematic pattern in the vowel alternations, they would simply take the surface representations as the underlying representation, as is in the case of Modern English. This is a traditional idea, tracing back at least to Kiparsky's Alternation Condition (1968), but now supplemented with a quantitative learning principle that specifies just how systematic an alternation needs to be to justify the postulation of abstract representations. We could also apply the method to, say, Old English, to see what would motivate a child learner to postulate an underlying representation (e.g., vowel length) that is distinct from its surface realizations. To do so, they must also acquire the morpho-phonological processes that manipulate and relate these (potential) representations. The reader is directed to Richter (2021) and Belth (2023) for interesting research pursuing these lines.

Our thought experiment has yielded some new insights and refined understanding of previous efforts. For one class of Middle English nouns,

i.e., those that have a schwa ending in the singular, they are mechanistically predicted to retain the long vowel with a tolerable number of exceptions according to the TP. The other nouns were predicted not to reach the requisite level of regularity for vowel length alternation, also according to the TP. For these, it seems that usage frequency has played an important but hardly deterministic role, as the threshold hypothesis reigns in the exaggerated effect of frequency in previous work (e.g., Bybee, 2010).

It is perhaps worth pointing out that all historical studies of languages are thought experiments. The language is dead and there are no native speakers around to consult. Nevertheless, linguists have been able to reconstruct the properties of numerous dead languages and their historical trajectories with astonishing detail and accuracy, based on the distributional properties of often fragmentary data. The same can be said about children. The linguistic experience for every child is a somewhat arbitrary sample of the language, which is, in turn, a somewhat arbitrary product of history. Yet children in the same speech community are able to acquire a largely uniform grammar (Labov, 2012). There must be a mechanism that reliably projects a grammar from children's messy data. Understanding this mechanism may help historical linguists understand their own messy data.

References

Allen, C. L. (1999). *Case marking and reanalysis. Grammatical relations from Old to Early Modern English*. Clarendon Press.

Árnason, K. (1998). Vowel shortness in Icelandic. In W. Kehrein & R. Wiese (Eds.), *Phonology and morphology of the Germanic languages* (pp. 3–25). De Gruyter Mouton.

Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.

Belth, C. (2023). Toward an algorithmic account of phonological rules and representations. Ph.D. dissertation. University of Michigan, Ann Arbor, MI.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Berman, R. A. (1981). Regularity vs anomaly. The acquisition of Hebrew inflectional morphology. *Journal of Child Language, 8*, 265–282.

Björnsdóttir, S. M. (2023). Predicting ineffability. Grammatical gender and noun pluralization in Icelandic. *Glossa. A Journal of General Linguistics,* 8. https://doi.org/10.16995/glossa.5823

Björnsdóttir, S. M. (2021). Productivity and the acquisition of gender. *Journal of Child Language, 48*, 1209–1234.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children. Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75*, 1115–1139.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis. A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425–455.

Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge University Press.

Caplan, S., Kodner, J., & Yang, C. (2020). Miller's monkey updated. Communicative efficiency and the statistics of words in natural language. *Cognition, 205*, 1044–1066.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. MIT Press.

Dresher, B. E., & Lahiri, A. (2022). The foot in the history of English. Challenges to metrical coherence. In Los et al. 2022: 42–59.

Emond, E., & Shi, R. (2021). Infants' rule generalization is governed by the Tolerance Principle. In D. Dionne & L.-A. V. Covas (Eds.), *Proceedings of the 45th annual Boston University Conference on Language Development* (pp. 191–204). Cascadilla Press.

Fenson, L., Dale, P. S., Steven Reznick, J., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). *Variability in early communicative development*. Society for Research in Child Development.

Gleitman, L. R., & Trueswell, J. C. (2020). Easy words. Reference resolution in a malevolent referent world. *Topics in Cognitive Science, 12*, 22–47.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*, 515–531.

Gorman, K., & Yang, C. (2019). When nobody wins. In F. Rainer, F. Gardani, H. C. Luschützky, & W. U. Dressler (Eds.), *Competition in inflection and word formation* (pp. 169–193). Springer.

Greenberg, J. H. (1966). *Language universals*. Mouton.

Halle, M. (1962). Phonology in generative grammar. *Word, 18*, 54–72.

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry, 4*, 3–16.
Henke, R. E. (2023). Rules and exceptions; A Tolerance Principle account of the possessive suffix in Northern East Cree. *Journal of Child Language, 50*(5), 1119–1154. https://doi.org/10.1017/S0305000922000277
Hogg, R. M. (1992). Phonology and morphology. In R. M. Hogg (Ed.), *The Cambridge history of the English language i, The beginnings to 1066* (pp. 67–167). Cambridge University Press.
Jakobson, R. (1932). Zur Struktur des russischen Verbums. In V. Mathesius (Ed.), *Charisteria mathesio oblata quinquagenario a discipulis et circuli linguistici pragensis sodalibus oblata* (pp. 74–83). Pražský Linguistický Kroužek.
Jakobson, R. (1971). *Selected writings i, Word and language*. Mouton.
Kiparsky, P. (1968). *How abstract is phonology?* Indiana University Linguistics Club.
Kodner, J. (2019). Estimating child linguistic experience from historical corpora. *Glossa, 4*, 122. https://doi.org/10.5334/gjgl.926
Kodner, J. (2023). What Latin verbal morphology tells us about morphological theory. *Natural Language and Linguistic Theory, 41*, 733–792.
Kroch, A., Santorini, B., & Delfs, L. (2004). *Penn-Helsinki Parsed Corpus of Early Modern English*.
Kroch, A., & Taylor, A. (2000). *Penn-Helsinki parsed corpus of Middle English*[2].
Kučera, H., & Nelson Francis, W. (1967). *Computational analysis of present-day American English*. Brown University Press.
Labov, W. (2012). What is to be learned. *Review of Cognitive Linguistics, 10*, 265–293.
Lahiri, A., & Elan Dresher, B. (1983). Diachronic and synchronic implications of declension shifts. *The Linguistic Review, 3*, 141–163.
Lahiri, A., & Elan Dresher, B. (1999). Open syllable lengthening in West Germanic. *Language, 75*, 678–719.
Lass, R. (1992). Phonology and morphology. In N. Blake (Ed.), *The Cambridge history of the English language ii, 1066–1476* (pp. 23–155). Cambridge University Press.
Lightfoot, D. W. (1979). *Principles of diachronic syntax*. Cambridge University Press.
MacWhinney, B. (2000). *The CHILDES project. Tools for analyzing talk*[3]. Lawrence Erlbaum.
Maratsos, M. (2000). More overregularizations after all. New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen and Xu. *Journal of Child Language, 27*, 183–212.

Marcus, G., Pinker, S., Ullman, M. T., Hollander, M., Rosen, J., & Fei, X. (1992). *Overregularization in language acquisition*. University of Chicago Press.

Masdottir, T. (2008). Phonological development and disorders in Icelandic-speaking children. Ph.D. dissertation, University of Newcastle upon Tyne.

Merkuur, A. (2021). Changes in modern Frisian verbal inflection. Ph.D. dissertation, University of Amsterdam.

Minkova, D. (1982). The environment for open syllable lengthening in Middle English. *Folia linguistica historica, 16*, 29–58.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304–330.

Nida, E. A. (1949). *Morphology. The descriptive analysis of words*[2]. University of Michigan Press.

Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Max Niemeye.

Pearl, L., & Sprouse, J. (2021). The acquisition of linking theories. A tolerance and sufficiency principle approach to deriving UTAH and rUTAH. *Language Acquisition, 28*, 294–325.

Plunkett, K., & Marchman, V. A. (1993). From rote learning to system building. Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.

Prokosch, E. (1939). *A comparative Germanic grammar*. Linguistic Society of America.

Richter, C. (2021). *Alternation-sensitive phoneme learning. Implications for children's development and language change*. Ph.D. dissertation, University of Pennsylvania.

Ringe, D., & Yang, C. (2022). The threshold of productivity and the irregularization of verbs in Early Modern English. In Los, B. et al. (Eds.), *English historical linguistics: Change in structure and meaning* (pp. 91–111). Amsterdam: John Benjamins.

Schuler, K. (2017). *The acquisition of productive rules in child and adult language learners*. Ph.D. dissertation, Georgetown University.

Schuler, K., Yang, C., & Newport, E. (2016). Testing the Tolerance Principle. Children form productive rules when it is more computationally efficient to do so. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th annual meeting of the cognitive science society* (pp. 2321–2326). Lawrence Erlbaum Associates.

Sneller, B., Fruehwald, J., & Yang, C. (2019). Using the tolerance principle to predict phonological change. *Language Variation and Change, 31*, 1–20.

Taylor, A., Warner, A., Pintzuk, S., Arja, N., & Nevalainen, T. (2006). *The parsed corpus of Early English correspondence*. University of York and University of Helsinki.

Thordardottir, E. T., Weismer, S. E., & Evans, J. L. (2002). Continuity in lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First Language, 22*, 3–28.

Tiersma, P. (1982). Local and general markedness. *Language, 58*, 832–849.

Trips, C., & Rainsford, T. (2022). Tolerating subject-experiencers? Yang's tolerance principle applied to psych verbs under contact in Middle English. *Journal of Historical Syntax, 6*, 1–43.

van Tuijl, R., & Coopmans, P. (2021). The productivity of Dutch diminutives. *Linguistics in the Netherlands, 38*, 128–143.

Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press. *Proceedings of the 45th annual Boston University Conference on Language Development*, ed.

Wright, J., & Wright, E. (1928). *An elementary Middle English grammar*. Oxford University Press.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook, 5*, 333–370.

Yang, C. (2016). *The price of linguistic productivity. How children learn to break rules of language*. MIT Press.

Yang, C. (2017). How to wake irregular (and speechless). In C. Bowern, L. Horn, & R. Zanuttini (Eds.), *On looking into words (and beyond). Structures, relations, analyses* (pp. 211–232). Language Science Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort. An introduction to human ecology*. Addison-Wesley.