

Metalinguistic, shmetalinguistic: the phonology of shm-reduplication

Andrew Nevins and Bert Vaux
MIT and Harvard University

1 Introduction

English shm-reduplication, as in *Oedipus shmoedipus*, is well known but incompletely understood. McCarthy and Prince (1986:71) for instance state that “the initial consonants, if any, are replaced by a specified consonant or cluster” [in this case *ʃm-*]. Sanders 2000 similarly states “a word is fully reduplicated, and the initial onset of the reduplicant is replaced with *schm-*”. Such analyses leave many subcases of shm-reduplication unexplained; what about cases where shm- does not replace the entire onset, as in *breakfast-shmreakfast* (in the film *Mall Rats*)? What about *obscene*, *obschmene*, where shm- overwrites a medial onset? What about *arcade*, *shmárcade*, where the stress of the base word changes?

Despite frequent references in the literature (e.g. Feinsilver 1961, McCarthy and Prince 1986, Zwicky and Pullum 1987, Pullum 1991, Alderete et al. 1996, Ginzburg and Sag 2000), English shm-reduplication and its attendant problems have yet to receive a thorough or satisfactory linguistic study. To address these lacunae we conducted a detailed online survey.¹ Based on the principled variation that emerged, we characterize a restricted set of morphophonological operations with implications for both rules and representations as they may variably target prosodic landmarks (first vs. stressed syllable), syllabic landmarks (onset vs. first consonant) and the phrasal size of the target of reduplicative morphemes (first word vs. entire phrase). Second, based on subjects’ treatment of *s-*, *m-*, *schm-*, and *sh-* initial words, and words whose output is an existing *schm-* initial word (e.g. *Joe schmo*), we argue that avoidance phenomena show a regularity that, contra Pullum 1991 and Sanders 2000, cannot be dismissed as metalinguistic.

2 How did shm-reduplication develop?

Though shm-reduplication is most familiar from English, individuals who are familiar with it generally feel it to be of Yiddish origin. Southern (forthcoming) suggests that shm-reduplication arose in Yiddish from a mix of Turkic Echo *m-* and East Slavic *sh-*. The Oxford English Dictionary on the other hand sees it as an English-internal development, “derived from the numerous Yiddish words that begin with this sequence of sounds”. The existence of early Yiddish forms in *shm-* supports the former theory over the latter (cf. Weinreich (1980:623-4), who seems to think that the construction goes back several centuries in Yiddish.) Southern cites in his support Yiddish *shmallig*, employed in a manuscript of c.1600 to disparage *hallig* ‘holy’. Yitskhok Niborski (personal communication)

¹ The survey can still be taken at <http://www.ai.mit.edu/projects/dm/shm>.

hypothesizes that the archetype for shm-reduplication in Yiddish is the collocation *tate shmate* ‘father shmather/rag’, which he states was already in use more than 150 years ago in European Yiddish communities. It would have been used, he adds, by an embittered wife against the man who provided her with children but not with an income. In this case *shmate* is an independent lexical item meaning ‘rag’, but it may have provided the vehicle for reanalysis as an echo formation.

Lockwood 1978 claims that shm-reduplication migrated into from Yiddish into American English in the late 19th century, and became integrated into common usage by the 1930s. The Oxford English Dictionary cites its first known use in English as *crisis-shmisis* (Goller 1929:V.ii.215).

Today shm-reduplication enjoys a fairly wide distribution in English, and is not limited to Yiddish or Jewish contexts, as shown by the following:

- (1) “Pedro Schmedro. The third-place Red Sox may have their big gun going to the mound this afternoon at the SkyDome, but as far as the exuberant Blue Jays are concerned, it’s a case of too little, too late.” (The Toronto Star, June 25, 2000)
- (2) “Breakfast?! Breakfast shmreakfast, look at the score for God’s sake. It’s only the second period and I’m winning twelve to two. Breakfasts come and go, Rene, but Hartford, the Whale, they only beat Vancouver maybe once or twice in a lifetime.” (Mallrats, 1995)

3 What does shm- mean?

Shm-reduplication resembles echo formations in other languages in being used to downplay or deride a particular phrase (cf. Emeneau 1939 for South Asian languages). As one survey respondent put it, applying shm-reduplication to a form indicates “I care so little about [it] that I will pronounce it flagrantly incorrectly, *so there*”. The dismissive sense of the construction can also be employed modally, to reassure, to downplay a situation or problem that is potentially overwhelming or threatening, or to lighten a situation with humor by pretending to dismiss it.

The elements of a shm-reduplication generally take the form of a topic-comment binomial pair, often with incomplete list intonation. As we have discussed with Michael Wagner, it may be that the intonation of shm-reduplication is crucial to deriving its meaning: the intonation of enumerating a list of items that share a common property, whose second item is in this case, a total nonsense word, may indicate that the first item resides in a natural class with unimportant items.

4 The survey

The semantics and pragmatics of shm-reduplication are relatively clear, but its phonological, morphological, and syntactic behavior are less so. In order to

elucidate the nature of these aspects of the phenomenon, we conducted an online survey consisting of 55 questions designed to elicit information on the treatment of variables including m-initial inputs, shm-initial inputs, non-initial shm, glide-initial inputs, second-syllable stress, ambisyllabicity, compounding, phrase structure of different types, lexical blocking, and complex onsets. The order of the questions and of answers was randomized for each survey respondent. Each question was multiple choice, in the following general format:

- (3) **Person A:** Who did that awful painting hanging in your basement?
Person B (pointing at her husband): Umm, Lee did.
Person A: Lee did, _____! I know that *you* painted it, you untalented fop!
 A. Lee shmid
 B. Shmee shmid
 C. Shmee did
 D. Nothing sounds good here
 E. Other []
 Comments on this question []

As of this writing, 190 individuals have completed the survey; our discussion and analysis are based on the results from these. Since some respondents did not complete the survey and the order of questions was randomized each time the survey was administered, each question has fewer than 190 data points.

We then administered a second survey of 200 individuals, to test additional variables and establish a baseline for shm-reduplication in basic cases.² All 200 individuals answered each of the five questions in this second survey.

5 Analysis

We begin our analysis with the phonological behavior of shm-reduplication. We situate our interpretation of the reduplicative and overwriting aspects of the phenomenon within the theory of Anchor Points developed by Yu 2002 and Nevins & Vaux 2003, which in brief hypothesizes the following:

- (4) Phonological Rules cannot count.
 (5) Phonological Rules can only refer to a restricted set of Anchor Points.

In this theory, processes such as infixation and reduplication can only target one of the following prominent landmarks (cf. Pierrehumbert and Nair 1995):

- (6) Anchor Points: 1st σ , 1st foot, 1st consonant, 1st vowel, stressed σ , final σ

² This survey can be seen at survey.net.nz/survey.php?7b71e88ecfa1f070b7199fd4f5fbf785.

Each of these targets is attested for both reduplication and infixation:

- (7) **Infixation after first foot:** Ulwa *karas:-ka-mak* ‘knee’ (McCarthy and Prince 1993)
- (8) **Reduplication after first foot:** Yidiny *mula-mula-ri* ‘initiated man’ (Dixon 1977, cited in McCarthy and Prince 1999)
- (9) **Infixation after Initial Consonant:** Sundanese *n-ar-aho* ‘to know’ (Robins 1959); Atayal *k-maial* ‘talk (actor focus)’ (Egerod 1965)
- (10) **Reduplication after Initial Consonant:** Mangarrayi *g-ab-abuji* ‘old persons’ (Downing 2002)
- (11) **Infixation after first vowel:** Katu *ka-r-chet* ‘dead’ (Costello 1988)
- (12) **Reduplication after first vowel:** Ø (but Pima may provide an example)
- (13) **Infixation Before Stressed σ :** English *e-goddamned-váporate* (McCarthy 1982)
- (14) **Reduplication Before Stressed σ :** Chamorro *hu-ga-gándo* ‘playing’ (Topping 1973)
- (15) **Infixation before final σ :** Hua *haru-ʔa-po* ‘not slip’ (Haiman 1980)
- (16) **Reduplication before final σ :** Dakota *spa-ha-ha* ‘broken off’ (Shaw 1976)

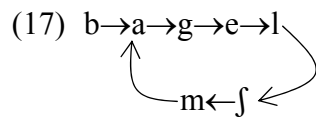
As observed in Raimy 2002:69-70, in a representational model in which reduplication is represented as an operation of introduction of precedence relations, infixation and reduplication become quite similar; both involve an anchor-point specification in the realization of a morphosyntactic feature. We claim that infixation and reduplication can only target this set of positions. Infixation is the introduction of a precedence relation (including new segmental material) whose start point transitively precedes its end point, while reduplication is the introduction of a precedence relation (including new segmental material in cases of fixed segmentism) whose start point is transitively preceded by its end point. Given their representational similarity, one expects that these two processes can target the same set of non-edge constituents. The logical conclusion for Yu is that infixation (and presumably reduplication as well) must be stated as inviolable, non-gradient Align constraints. We adopt this inventory of anchor points (edge constituents and prominent constituents) as complete, noting that restricting precedence-introducing relations to these points prevents the grammar from “counting”, and disables infixation or reduplication after, say, the third vowel in an octosyllabic word.³

We hypothesize that the set of anchor points provided by the above theory accounts for the range of attested variation in shm-reduplication. That is, speakers

³ Thus mitigating the concerns raised by Downing 2001 about the generative space of precedence-based reduplication.

will vary as to where they insert *shm-*, but the range of variation will be confined to the inventory of anchor points in (6). When a learner hears *bagel shmagemel*, there are many hypotheses compatible with this single pattern (overwrite the first consonant, overwrite the first onset, overwrite up to the nucleus, overwrite in the stressed syllable, etc.). The underdetermined nature of this sort of data allows speakers to postulate a range of rule targets, all of which are compatible with the data for simple (i.e. one-consonant-onset with initial stress) forms, but reveal variation when speakers are required to produce more complex forms.

Turning next to mechanics, the fundamental operations for *shm-reduplication* consist of three steps: (i) adding a new precedence relation from the last segment of the input word to the sequence /ʃm-/, (ii) adding a new precedence relation from the sequence /ʃm-/ to the anchor point (i.e. postulated target) of the rule, and (iii) linearizing the structure. Step (ii) is the locus of variation. The most common target of (ii) in *shm-reduplication* is the first vowel. Thus, where \rightarrow denotes immediate precedence, most speakers postulate the representation as in (17):



The representation in (17) is within the precedence-based model of Raimy 2000 (see also Fitzpatrick & Nevins 2003). We adopt a precedence-based model because it allows explicit formalization of the target of a morphophonological rule resulting in reduplication. In this model, the output token [bagel-shmagel] is phonologically *ambiguous*, in the way that *I saw the man with the telescope* is—it has more than one possible structural analysis. Generalization of the phonological realization of the semantic target of *shm-reduplication* from one of these analyses of the ambiguous string generates in the massive variation we document in this paper, as exemplified in (18).

- (18) variation by anchor point in *shm-reduplication*
- a. targets **first syllable** versus **stressed syllable**
shmobsce (30% of respondents) vs. *obshmene* (28%)⁴
 - b. *shm* placed at target in stressed syllable versus second copy **starts with shm at target** in stressed syllable
confusion conshmusion (34%) vs. *confusion s(h)musion* (11%)⁵
 - c. targets material after **first consonant** versus **first onset**
breakfast s(h)mreakfast (10%) vs. *breakfast s(h)meakfast* (87%)

⁴ 47/128 respondents (37%) preferred no output for *obsce*; 1 respondent favored *shmene*.

⁵ 47/113 respondents (42%) preferred no output for *confusion*; 13% preferred *shmonfusion*.

For the purposes of comparison, we will address why we do not adopt the shm-reduplication analysis of Alderete et. al 1999. In their model, *bagel*, *shm-* and RED (the abstract instruction to reduplicate) are inputs to a computation that is based on violation counting. Violations are counted such that it is worse for the output to miss something in the input (MAXIO) than it is for the reduplicant to contain something that is not in the base (MAXBR); i.e. MAXIO >> MAXBR. The output [bagel-shmbagel] would satisfy both of these constraints, but is rejected for phonotactic reasons. In the computation of [bagel-shmagel] vs. [shmagel-shmagel], the latter loses because it fails to include [b] from the input. However, with this same computation, the optimal output for the input /eel/ is *[shmeel-shmeel]. Moreover, while the analysis predicts [wig-shmig] as optimal, 10/115 of our respondents did not choose this option. The issue of how to deal with all of the variation that we successfully model in the current paper is not addressed, as Alderete et. al idealize to a unique output in English for shm-reduplicating *bagel*.

In addition, the crosslinguistic generality of this model is questionable, as Hindi v-reduplication yields [roti-voti] and never *[roti-vroti] for the input /roti/, even though the latter output should be a penalty-minimization dream come true. Finally, for proponents of the view that a simple re-ranking of the constraints MAXIO >> MAXBR should produce an attested reduplication pattern, it is demonstrable that their re-ranking will yield a reduplicative system in which an input with *three* onset consonants will never deliver shm- reduplication, and the optimal output for /string/ would be [string-string]. As is more fully discussed in Nevins 2003, these properties of Alderete et al.'s model all result from the fact that computation is done in terms of *counting* the number of segments in violation; in this case, because 3 is more than 2. Our model, as mentioned above, is exempt from counting, and avoids such pathological predictions.

5.1 Complex Onsets

With these theoretical preliminaries in mind, we now turn to the data. One of the first variations that we noticed in shm-reduplication is between those who do and do not preserve elements of word-initial consonant clusters. In order to determine the parameters of this variation, we included in our survey four questions containing (or potentially producing) different types of word-initial clusters: Cr (*breakfast*), CCr (*street*), Cl (*floozie*), and r (*rich*). We also included two stimuli of this type in the follow-up survey, *broom* and *floss*.

The treatment of Cr clusters was evenly split between the two options mentioned above when no other factors interfered, as can be seen for *broom* in (19).

- | | | | |
|------|----|---------------|-----|
| (19) | a. | broom shmoom | 39% |
| | b. | broom shmroom | 37% |
| | c. | broom smroom | 6% |
| | e. | no output | 18% |

The remaining Cr stimuli showed a greater preference for the r-less treatment:

(20)		with r	without r	no output
a.	breakfast	10%	87%	3%
b.	street	4%	71%	22%
c.	rich	3%	82%	13%

Cl clusters also display preference for the l-less treatment:

(21)		with l	without l	no output
a.	floss	16%	72%	11%
b.	flooze	3%	88%	10%

The CR- cluster data in (19-21) suggest that speakers are most likely to form the strongest generalization—target the segment immediately following the first onset—when presented with underdetermined data of the *bagel shmager* type. Note, though, that a non-trivial percentage of speakers select the more restrictive hypothesis that targets the segment immediately following the first consonant.

5.2 Glides

Recently, several phonologists have debated an issue related to the complex onset problem just discussed, the behavior of post-consonantal glides. Barlow 2001 and Yip 2003 have suggested on the basis of Pig Latin that English speakers may have two representations for a sequence like [spju], one that syllabifies the [j] in the onset, and one that attaches it to the nucleus. Nevins & Vaux 2003 and Idsardi & Raimy 2003, on the other hand, have suggested that English post-consonantal [j] is always part of the nucleus, and that the attested variation in reduplicative behavior results from two different rule variables.

To investigate this issue we tested five words with appropriately situated post-consonantal glides. Our results are tabulated in (22), where “+glide” is a cover term for forms like *union shmyoonion* and *wig shmwig*, and “-glide” represents forms that do not copy the glide, such as *union shmoonion* and *wig shmig*.

(22)	base form	+glide	-glide	null output	total responses
a.	union	19%	75%	5%	114
b.	use (verb)	8%	68%	22%	116
c.	confusion	13%	38%	48%	98 ⁶
d.	wig	5%	92%	3%	114
e.	dwarf	31%	63%	6%	199

⁶ 15 additional respondents said *shmonfusion*, which is not relevant to the treatment of post-consonantal glides. These were not included in the computation.

We account for the variation in (22) by postulating that speakers hypothesize two different targets for the reduplicative pointer: one group adds it from /ʃm/ to the first nuclear segment, giving e.g. *union shmyoonion*, and the other adds a relation from shm- to the first vowel, yielding e.g. *union shmoonion*. Speaker variation thus depends on the target of the reduplicative pointer, not the representation of glides. This analysis has the advantage over Barlow and Yip's theory of employing a uniform representation for all prevocalic glides, and explaining the fact that speakers may treat /ju/ sequences differently in different language games (cf. Idsardi & Raimy 2003)—the glide as a prenuclear segment remains constant, but the game-particular rule contains a different anchor point variable.

5.3 [+anterior] dissimilation

Having encountered individuals who employ sm- rather than shm- with words containing a strident consonant {ʃ tʃ ʒ dʒ s z}, we included in the survey several questions to test the extent of this phenomenon. Our findings are summarized in (23), where “ʃm” is a cover term for forms like *witches shmitches*, and “sm” represents forms that employ sm-, such as *witches smitches*.

(23) base form	ʃm	sm	null output	total responses
a. witches	70%	17%	12%	114
b. Ashmont	70%	17%	13%	115
c. ash	66%	10% ⁷	21%	115
d. massage	43% ⁸	4%	48%	122
e. schnozz	66%	3%	27%	113

Why does ʃ change to s precisely in forms containing one or more stridents? We propose that it results from the rule of [+anterior] dissimilation represented in (24):

(24)	[+strident]	[+strident]
	X	X
	≠	
	[-ant]	[αant]

The system that produces *ash shmass* modifies (24) by delinking the second [+anterior] specification rather than the first.

5.4 Non-initial stress

Most varieties of shm-reduplication require initial stress on both the base and the reduplicant. What happens when the base does not have initial stress? Canonical

⁷ 10 respondents produced *smash*, and 2 produced *shmass*.

⁸ Responses grouped in this category were *shmassage* (39), *shmage* (7), and *mashmage* (6).

cases of shm-reduplication do not enable us to answer this question, because they all involve bases with primary initial stress. It turns out that speakers employ a variety of different strategies: if the initial syllable bears a secondary stress most speakers promote it, as in *économics shméconomics*; if it does not, as in *arcáde*, speakers split fairly evenly between preserving the base stress in the reduplicant (*arcáde shmarcáde*, 40%) and overriding the base stress (*árcade shmárcade*, 47%). The alteration of the base stress in the latter form is noteworthy; it is not clear whether this is due to the rhythm rule or a desire for prosodic identity between the based and the reduplicant.

We tested for potential influence of lexical competitors by including the verb *perμίt*, which contrasts with its nominal counterpart only in terms of stress, the former being final and the latter being initial (*pérμίt*). Of the 107 individuals who answered this question, only 12 (11%) allowed the verb to shm-reduplicate, whereas 105 (98%) allowed shm-reduplication of the noun.

We also discovered that it is not uncommon to shm-reduplicate internally on the stressed syllable in words with non-initial stress, as shown in (25):

(25)	form	initial	%	internal	%	null
a.	confusion	shmonfusion	13	(con)s(h)musion ⁹	44	42%
b.	obscene	shmobsceñe	33	(ob)shmene ¹⁰	31	40%
c.	massage	s(h)massage	36	(ma)s(h)mage ¹¹	11	48%
d.	terrific	shmerrific	67	tershmiffic ¹²	5	26%
e.	arcade	shmarcade	87	(ar)shmade ¹³	3	9%

Though not evident from the forms in (25), internal shm-reduplication is not restricted to the second syllable; we have heard forms such as *regulations*, *regushmations* and *understandable*, *undershmandable*. Since in the latter case the stress coincides with a morpheme boundary, we decided to include a form where the two do not align, *unbelievable*, and found that 45% produced *shmunbelievable*, 38% preferred no output, 7% chose *unshmelievable*, 6% picked *unbeshmievable*, and 3% produced *shmievable*.

The last of these options shows not only that shm-reduplication for some speakers refers to the stressed syllable, but also that the reduplicant may begin from this

⁹ 34% produced *conshm(y)ooision*, 8% produced *shmoosion*, and 3% produced *smoosion*.

¹⁰ 36 tokens of *obshmene*, and 1 of *shmene*.

¹¹ 6 respondents gave *mashmage*; 7 gave *shmage*.

¹² Note the preservation of the base /r/, suggesting that it is ambisyllabic for these speakers.

¹³ 2 respondents produced *arshmade*, and one produced *shmade*.

syllable. Such forms have been produced by Spitzer 1952 (*confusion shmoooshun*) and Howard Stern (*massage shmage*), among others.

Interestingly, a small number of individuals associate shm- to both the initial and the stressed syllables, producing forms such as *forbidden*, *shmorshmidden*, which appeared in an episode of the Fox program *Futurama*.

6 Morphology and syntax

Having examined the phonological variation that results from the inventory of anchor points and the nature of phonological rule postulation, we turn to variation in another domain: morphologically complex words. As it is impossible that shm-reduplication happens “in the lexicon”, that leaves one place for word-formation to happen: in the syntax. Though shm- reduplication has a phonological form that looks a bit different from more concatenative modes of morpheme exponence, its composition in the syntax is still achieved through the operation of Merge. The source of variation that we observe arises from the fact that the morpheme of which shm-reduplication is an exponent (call it ECHO, which is quite likely featurally underspecified and derives its interpretation through pragmatics) may attach at multiple levels in the syntactic tree.

As Lidz (2001) discusses, echo reduplication can apply at various points within a syntactic structure. For example, Kannada gi-reduplication can occur either inside or outside of the case-marker on a noun. Thus, the input [baagil] ‘door’ can be merged with ECHO and then [[baagil] ECHO] can be merged with the K(ase) morpheme -annu (26a), or [baagil] can be merged with K and then [[baagil] K] can be merged with ECHO (26b).

- (26) a. *baagil-giigil-annu* ‘doors-shmoors-acc.’
- b. *baagil-annu giigil-annu* ‘doors-acc-shmoors-acc’

There is no semantic difference between these derivations; it constitutes a straightforward case of two sequences of Merge with the same morpheme inventory, resulting in radical variation once the structure is phonologically spelled out.

The results of our survey demonstrate that variation in the site of merge of ECHO in the syntax obtains in shm-reduplication as well. For instance, given a phrasal input, our respondents varied as to whether shm-reduplication applied to the first XP in the phrase (27b-c), or the most deeply embedded (27a):

- (27) Q16 **Person A:** Dear, you'd better stop squandering our money on golf clubs. Jacob wants a laptop, and you owe him big-time after embarrassing him in front of his friends by getting Chumbawumba to play at his bar mitzvah.

Person B: Jacob wants a laptop, _____! That spoiled brat has enough junk already, and Chumbawumba were ahead of their time!

- a. shm- targets *laptop* (9%)
- b. shm- targets *Jacob* (5%)
- c. shm- targets both *Jacob* and *laptop* (3%)
- d. Ineffable for whole sentences (78%)

We also examined shm-reduplication in endocentric compounds:

(28) Q11 **Person A:** You know what we really need? A cookie cutter.

Person B: Cookie cutter, _____! We have enough kitchen utensils!

- a. shm- targets *cookie* (65%)
- b. shm- targets *cutter* (10%)
- c. shm- targets both *cookie* and *cutter* (7%)
- d. Ineffable for compounds of this type (19%)

It is notable that exocentric compounds (e.g. *walkman*) are significantly more resistant to internal schm-reduplication (contrast (29a-b)), perhaps indicating that these are not built by the syntax.¹⁴

- (29) a. shm- targets *man* (1%)
- b. shm- targets *walk* (93%)
- c. shm- targets both *walk* and *man* (3%)
- d. Ineffable for compounds of this type (3%)

The variation in (28), however, clearly demonstrates that ECHO can attach in various places in compounds, and supports Lidz's (2001) hypothesis that morphological optionality is the result of different Merge sites.

In an interesting paper, Travis 2001 proposes that syntactic reduplication (e.g. *I need a DOCTOR doctor*) involves movement of a copy to check a feature. Presumably, one copy of the XP stays in base position, while a higher one moves to the head with the uninterpretable feature. However, we would like to point out that shm-reduplication can *never* appear in argument position (in fact, in stark contrast to e.g. Hindi v- echo reduplication).

- (30) a. *I don't want to go to Europe, Shmeurope
- b. Europe, Shmeurope, who wants to go there!

(31) Optimality, schmoptimality—who needs *(it)?

¹⁴ Names, on the other hand, seem to allow a variation that we have little explanation for in syntactic terms, e.g. *Donald Shmumsfeld* (26.5%), *Shmonald Rumsfeld* (18.5%).

We suggest that shm-reduplication involves obligatory movement to a Topic phrase (above the position of wh-movement, as shown by (30)) from an argument position, often with obligatory resumption (as in (31)) in the base position.

7 Coprecedence, and allomorphy

Shm-reduplication has received some discussion as an example of an avoidance phenomenon. McCarthy and Prince 1986:68 observe that “even English speakers with little experience of the phenomenon...report that words already beginning with the cluster *shm* cannot enter into this pattern: **shmaltz-shmaltz* (with the intended reading).” They then claim that speakers confronted with shm-initial words replace the overwriting -m- with -p- (*shmaltz shpaltz*), but our study reveals a far greater range of variants, as shown in (32) and (33):

- | | | | | |
|------|-------------|---------------|----------------------|-------------------------------------|
| (32) | base | % shm- | % null output | % other overwriting material |
| | a. Schmidt | 3 | 76 | 21 |
| | b. schmooze | 5 | 64 | 31 |
| | c. schmuck | 5 | 67 | 28 |
- (33) alternative overwriting sequences
- schmuck*: shluck, shnuck, schnook, fluck, shpuck, vluck, shmluck, shuck, shfuck, shvuck, smuck, fuck, shmukster, my ass, (Bronx cheer)
 - schmooze*: shnooze, flooze, shmooze, shpooze, shlooze, vlooze, shplooze, shməmooze, mooze, wooze, commooze, my ass

Some have suggested that shm-reduplication is infelicitous with shm-initial words because the latter already contain the derisive semantic connotations shm-reduplication is intended to produce. This hypothesis fails to explain the fact that words lacking these semantics, such as *Schmidt*, show the same avoidance effects.

We propose instead to link this phenomenon to the cross-linguistic fact about most (but not all—cf. Bengali *t-*) instances of echo reduplication that when the fixed segmentism would result in identical base and reduplicant, avoidance occurs, as exemplified in (34)-(36).

(34) Hindi *aam vaam, paani vaani, wakil fakil*

(35) Turkish *kitap mitap, *masa masa*

(36) Turkish *cip ciliz, dop doluz, dim dik* (dissimilation with obstruents; cf. Kelepir 1999)

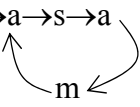
A blanket statement that Antifaithfulness or non-identity is required does not explain why it is precisely the fixed segment that dissimilates: a presumable solution to an Antifaithfulness requirement for shm-reduplication would be *schmear*, *schmeak*¹⁵, but Turkish, Hindi, and English speakers always dissimilate the fixed segment, as we saw in (34).

A theoretical question that arises, however, is how long-distance dissimilation of a potentially unbounded sort can be stated in local terms. In fact, as Fitzpatrick and Nevins 2002 shows, the precedence-based model offers a quite natural derived relation, coprecedence, defined in (37).

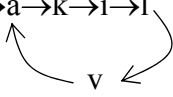
- (37) When $A \rightarrow C$ and $B \rightarrow C$, then A and B are *coprecedents*
(denoted by $A \oslash B$)

The environment for dissimilation is exemplified for Turkish (38) and Hindi (39):

- (38) Turkish m-reduplication (cf. (35))

- a. $m \rightarrow a \rightarrow s \rightarrow a$

 b. Ineffable structure: $m \oslash m$

- (39) Hindi Representative Plurality (cf. (34))

- a. $v \rightarrow a \rightarrow k \rightarrow i \rightarrow l$

 b. Dissimilation rule: $v \rightarrow \int / v \oslash _$

The statement of the rule for shm-reduplication is less systematic than it is in Hindi. While the structural change clearly differs for our respondents, the structural description for the rule, delivered by the formalism in (37), is virtually invariant: dissimilate when two $\int m$ sequences *coprecede*.

Avoidance allomorphy looks like long-distance dissimilation. But when two v 's are coprecedent this is a *local* relationship, and provides a context for dissimilation. Shm-dissimilation occurs not as metalinguistic correction (pace

¹⁵ To carry forth the spirit of the penalty-minimization view to its logical conclusion, we put the epenthetic segment at the end of the word, minimally disrupting continuity, and choose /k/ as the epenthetic segment, since we know, from the KEKU theorem (Fitzpatrick, Nevins, and Vaux 2003), that /k/ is as good a candidate for the most unmarked consonant as /t/.

Pullum 1991, Sanders 2000), but due to a simple coprecedence-based allomorphy rule.

8 Ineffability

The ineffability that we observe in many sub-cases of shm-reduplication, most notably with shm-initial words, raises another serious theoretical issue. As Gerfen 2001 observes, “a fundamental axiom of OT is that all constraints are violable”. The idea behind this generalization is that a grammar always picks the best of a set of candidate outputs, an idea that stands in explicit and intentional contrast to the inviolable constraints employed in pre-OT phonology. Absolute ungrammaticality directly contravenes OT’s central principle of Violability. Absolute ungrammaticality moreover plays a central role in human language, appearing robustly in all known phonological systems.

Prince and Smolensky 1993 propose to generate certain cases of ineffability by combining a Null Parse candidate, which is mysteriously stipulated to satisfy all well-formedness and faithfulness constraints, with a constraint MPARSE, which assigns a violation only to the Null Parse candidate. This analysis fails to capture intuition that (for example) shm-reduplication of *schmuck* produces *no* output, rather than a *phonetically null* output. Orgun and Sprouse 1999 demonstrate moreover that the Null Parse analysis creates unresolvable ordering paradoxes in Turkish. They conclude that grammars must be able to contain inviolable constraints, which requires that we abandon the pivotal OT notions of Violability and Emergence of the Unmarked. In contrast, the theory we employ in this paper directly accounts for ineffability in terms of inviolable constraints.

But what exactly motivates the inviolable constraint involved in echo reduplication? We tentatively suggest that nonrecoverability is involved: if one pronounces a form such as *schmidt schmidt*, *luck shmuck*, or *jews schmooze* (the name of Rob Reiner’s talk show in the film *Primary Colors*) it may not be sufficiently clear that shm-reduplication has occurred. This analysis may be able to help explain the different degrees of lexical blocking observed between forms where the output competes with a well-known word, such as *joe shmoe* (cf. *schmo*), and forms where the competing word is not well-known, as in *ear shmear* (cf. *shmeer*); the former case was ineffable for 34% of respondents, whereas the latter was only so for 10% of respondents.

Contributions

We have shown, through our survey, that speakers have clear and consistent linguistic intuitions, suggesting that shm-reduplication is computed in the grammar, and that the systematic responses of speakers to these forms show that this is not a metalinguistic phenomenon to be dismissed, and that the notion of metalinguistic is vacuous here, as speakers’ strategies still manipulate phonological objects. We have suggested that an individual’s grammar for shm-

reduplication is generally formed through underdetermined exposure, and we have identified several distinct subtypes of shm- reduplication, and shown that, as the variation is principled, it supports a theory of anchor points, as developed in Yu 2002 and Nevins and Vaux 2003.

Having demonstrated the importance of shm-reduplication as a source of data for models of phonological representation, word formation, and the status of ineffability in a grammatical computation, we would like to point out that the importance of shm-reduplication extends even beyond its illustration of the points we have discussed here. At this same conference, exactly twenty years ago, Alexis Manaster-Ramer demonstrated that the copying dependency in what he called “reshmuplication” constituted one of the first bona fide examples that the status of natural language within a complexity hierarchy could be conclusively derided as *context-free*, *context-shm(r)ee*.

Acknowledgements

We would like to thank all of the individuals who took our survey, and Toby Ayer, Sarah Benor, George Bergman, Susan Brunka, Justin Cooper, David Diaz, Noam Elkies, Elissa Flagg, Rebecca Starr, and Donca Steriade for helpful comments on earlier versions of this paper. Special thanks to Charlie Cheever, Marc Prud’hommeaux, Jeff Surette, and Dorothy Weiss for help with designing the survey and tabulating and analysing the data.

References

- Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy, and Suzanne Urbanczyk. 1996. Reduplication and Segmental Unmarkedness. Manuscript, University of Massachusetts, Amherst. [Different than the version later published in LI, which does not include the section on schm-reduplication. Available as ROA 134-0996.]
- Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy, and Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30.3: 327-365.
- Bagemihl, Bruce. 1988. *Alternate Phonologies and Morphologies*. Doctoral dissertation, University of British Columbia.
- Barlow, Jessica. 2001. Individual differences in the production of initial consonant sequences in Pig Latin. *Lingua* 111:667-696.
- Bruening, Benjamin. 1997. Abkhaz-Mabkhaz: M-Reduplication in Abkhaz, Weightless Syllables, and Base-Reduplicant Correspondence. In B. Bruening, M. McGinnis, and Y. Kang (Eds.), *PF: Papers at the Interface*. MIT Working Papers in Linguistics vol. 30. Cambridge, MA: MIT Working Papers in Linguistics.
- Costello, Nancy A. 1998. Affixes in Katu of the Lao P. D. R. *Mon-Khmer Studies* 28: 31-42.
- Dixon, Robert. 1977. *A grammar of Yidiny*. Cambridge: Cambridge University Press.
- Egerod, S. 1965. Verb inflection in Atayal. *Lingua* 15: 251-282.

- Feinsilver, Lillian. 1961. On Yiddish Shm-. *American Speech* 36: 302-3.
- Fitzpatrick, Justin and Andrew Nevins. 2002. Phonological Occurrences: Relations and Copying. Paper presented at NAPhC 2.
- Fitzpatrick, Justin, Andrew Nevins, and Bert Vaux. 2003. KEKU: On relative privilege of data in assessing relative markedness. Manuscript, MIT.
- Ginzburg, Jonathan and Ivan Sag. 2000. Interrogative Investigations: the form, meaning and use of English interrogative constructions. Stanford: CSLI.
- Goller, I. 1929. *Five Books of Mr. Moses*.
- Haiman, John. 1980. *Hua: A Papuan Language of the Eastern Highlands of New Guinea*. Amsterdam: John Benjamins.
- Lidz, Jeff. 2001. Echo Reduplication in Kannada and the theory of word formation. *The Linguistic Review* 18: 375-394.
- Manaster-Ramer, Alexis. 1983. The soft formal underbelly of theoretical syntax. *CLS* 19:256-262.
- McCarthy, John. 1982. Prosodic Structure and Expletive Infixation. *Language* 58.3: 574-590.
- McCarthy, John and Alan Prince. 1986. Prosodic Morphology 1986. <http://ling.rutgers.edu/gamma/pm86all.pdf>.
- McCarthy, John and Alan Prince. 1993. Generalized alignment. In Geert Booij and Jaap van Marle (eds.) *Yearbook of morphology 1993*, 79-153. Dordrecht: Kluwer Academics.
- McCarthy, John and Alan Prince. 1999. Faithfulness and Reduplicative Identity. In Kager, René, Harry van der Hulst, and Wim Zonneveld, eds., *The prosody-morphology interface*. Cambridge: Cambridge University Press.
- Nevins, Andrew. 2003. Overwriting: All or Nothing? Manuscript, MIT.
- Nevins, Andrew and Bert Vaux. 2003. Underdetermination in language games: Survey analysis of Pig Latin dialects. Paper presented at the LSA.
- Orgun, Cemil Orhan and Ronald Sprouse. 1999. From MPARSE to CONTROL: deriving ungrammaticality. *Phonology* 16.2: 191-224.
- Pullum, Geoffrey. 1991. *The great Eskimo vocabulary hoax, and other irreverent essays on the study of language*. Chicago: University of Chicago Press.
- Rackowski, Andrea. 1999. Morphological Optionality in Tagalog Aspectual Reduplication. *Papers on Morphology and Syntax, Cycle Two, MITWPL* 34: 107-136.
- Raimy, Eric. 2000. *The Phonology and Morphology of Reduplication*. Berlin: Mouton de Gruyter.
- Raimy, Eric. 2000. Remarks on backcopying. *Linguistic Inquiry* 31: 541-552.
- Robins, R. 1959. Nominal and verbal derivation in Sundanese. *Lingua* 8.

- Sanders, Nathan. 2000. Intra-Representational Correspondence and the Realization of Empty Morphemes. Manuscript, UC Santa Cruz.
- Shaw, Pat. 1976. *Theoretical issues in Dakota phonology and morphology*. Doctoral dissertation, University of Toronto.
- Southern, Mark. forthcoming. *Contagious Couplings: Yiddish shm- and the contact-driven transmission of expressives*. Westport, Conn.: Greenwood Publishing Group.
- Spitzer, Leo. 1952. Confusion shmooshun. *Journal of English and Germanic Philology* 51: 226-33.
- Steinmetz, Sol. 1986. *A Century of Yiddish in America: Yiddish and English*.
- Topping, Donald. 1973. *Chamorro Reference Grammar*. Honolulu: University of Hawaii Press.
- Travis, Lisa. 2001. The syntax of reduplication. In Minjoo Kim and Uri Strauss, eds., *Proceedings of the North Eastern Linguistics Society Annual Meeting (NELS 31)*, 455-469. Amherst, Mass.: GLSA.
- Weinreich, Max. 1980. *History of the Yiddish language*. Chicago: University of Chicago Press.
- Yip, Moira. 2003. Casting doubt on the Onset/Rime distinction. To appear in *Lingua*.
- Yu, Alan. 2002. Understanding infixes as infixes. Paper presented at the Second North American Phonology Conference, Montreal.
- Zwicky, Arnold and Geoff Pullum. 1987. Plain morphology and expressive morphology. *Proceedings of the Berkeley Linguistic Society*.