

The Coinages in Seuss*

Bruce Hayes
UCLA

August 2021

Submitted to *English Language and Linguistics*

Abstract

The children's books of Dr. Seuss abound in words that the author invented. Inspection shows that these coinages are not arbitrary, raising the challenge of specifying the linguistic basis on which they were created. Drawing evidence from regression analyses covering the full set of Seuss coinages, I note several patterns, which include coinages that are phonotactically ill-formed, coinages meant to sound German, and coinages that assist compliance with the meter. But the primary coinage principle for Seuss appears to have been to use words that include **phonesthemes** (Firth 1930), small quasi-morphemic sequences affiliated with vague meanings. For instance, the coinage *Snumm* contains two phonesthemes identified in earlier research, [sn-] and [-ʌm]. Concerning phonesthemes in general, I assert their affiliation with vernacular style, and suggest that phonesthemes can be identified in words purely from their stylistic effect, even when the affiliated meaning is absent. This is true, I argue, both for Seuss's coinages and for the existing vocabulary.

*This article has benefited from input by audience members in the UCLA Phonology Seminar and a later public lecture given by the author celebrating the creation of the Theresa M. and Henry P. Biggs Centennial Term Chair in Linguistics at UCLA. I would like to thank the Biggses for making the latter occasion possible. I would also like to acknowledge two persons who shared with me the exhilarating experience of reading Seuss aloud: my late mother, who read the books to me when I was a child, and my son Peter, to whom I read them when he was a child.

Since I have been working on Seuss, people occasionally have asked me about the political controversy currently surrounding him. My reply is to suggest reading about him before arriving at any conclusions. Morgan and Morgan (1995) is a good place to start.

1. Introduction: coined words

The reader of children's books by Dr. Seuss (Theodore Seuss Geisel, 1904-1991) cannot help but notice the great number of words Seuss coined himself. In (1) I give some examples, specifically the full set of coined words in *If I Ran the Circus* (1956).

(1) *The Seuss coinages in If I Ran the Circus*

Proper names:	<i>Sneelock, Bopps, Jawks</i>
Animals:	<i>Snumm, Foon, Walloo, Huffle, Fibbel, Snarp, Nolster</i>
Places and adjectives of origin:	<i>Jorn, Olf, Frumm, Brigger-ba-Root, Ben-Deezing, Soobrian</i>
Noises:	<i>bloop, floop</i>
Machines:	<i>bloozer, noozer</i>
Verbs:	<i>zoop</i>

Here is a passage employing two of these coined words:

- (2) From a country called Frumm comes this Drum-Tummied Snumm
 Who can drum any tune that you might care to hum.
 Doesn't hurt him a bit, cause his Drum-Tummy's numb.

In coining words, Seuss was hardly alone among authors of fiction; more exalted figures of English literature such as Jonathan Swift or James Joyce did the same. However, words are also coined by ordinary people from time to time (Marchand 1969: 320-321; Malkiel 1990:105). When adopted generally, these words end up in the dictionary, listed as words of obscure origin. Thus, for instance, Eisiminger (1981) compiled a list of English words that have no etymology, and it abounds in slangy, obviously-new forms. The same is true for older forms that gradually lost their slangy tinge and settled into standard usage (see, e.g. the *Oxford English Dictionary's* entries for *boy*, *girl*, *big*, and *bad*). So research questions about word coinage are not confined to literature but are part of the study of language in general. A prolific word-coiner like Seuss can help us to explore word coinage in one "idiolect" for which ample data are attested.

The freedom available to word-coiners is in principle very broad: it seems that they need only find *some* string that conforms to the phonotactics of their language enough to be pronounceable by other speakers. But this greatly oversimplifies the question. Word-coiners create their words for a reason, and they make substantial use of the phonological resources of their language when they create a novel phonological form. This point is well established, I believe, by recent research on word coinage, notably the extensive current research program on the creation of Pokémon names (for overviews, see Shih et al. 2018 and Kawahara 2021a). However, we will see that the Seuss corpus has idiosyncrasies that justify a slightly different analytical approach.¹

¹ The work just cited tends to track very general properties, such as "number of voiced obstruents" or "number of syllables," whereas I have found it useful to zero in on particular phonemic sequences. In Fordyce's

Before plunging into the Seuss coinages, I should offer a couple of clarifications. First, when I say “coinage” here, I obviously am using it in a restricted sense, namely “made up *de novo*”, since we can also say that word creation in the normal way — application of the language’s word formation rules, as in *Seuss-ian* or *un-mute*, counts as coinage. Second, I acknowledge that word coinage often relies on lexical as well as phonological resources: a coined word is sometimes perceived to be similar in its phonology and semantics to an existing word (MacDonald 1999:67; Magnus 2001:140). For simplicity, the discussion below will ignore this factor.

2. The Seuss coinages: an attempt at precise description

I adopt the strategy of Shih and Kawahara, employing a digital data corpus and statistical modeling in order to obtain objective testimony about issues that can become subjective very easily. The modeling described here employs the technique of *logistic regression* (on which see, e.g., Johnson 2008:159-174). The purpose of my models is to predict for any given word, on the basis of just its phonological form, whether it is a Seuss coinage or a real word (for similar applications in other domains, see Hayes 2016). As we will see, it is impossible to achieve perfect prediction, and indeed what the model really does is to assign a “probability-of-Seussian” value to every form, so its predictions are gradient. The deeper purpose of the modeling is that, once a model has been optimized, we can make useful inferences from its internal structure, specifically the degree to which the model attributes explanatory importance to principles hypothesized to underlie Seuss’s coinage practice.

I employed readily accessible data. The Seuss coinages, which number about 435 in his complete oeuvre, were carefully collected and described by Latham (2000). I extracted the coinages from Latham’s work and rendered them in phonetic transcription by hand. I believe the latter task is not difficult or controversial, in light of Seuss’s clear use of orthography and the additional clues provided by rhyme. For English in general I employed my own version of the Carnegie-Mellon pronunciation dictionary (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>). My edited version, with 17,744 entries, includes only words that have a frequency of one or above in the English CELEX database (Baayen et al. 1995). This is meant to restrict it to words likely to be familiar to English speakers. I also excluded words formed with highly productive suffixes such as inflectional [-z/-s/-əz] (plurals, possessive, 3rd sg. pres.) or [-d/-t/-əd] (past tenses and participles). This is important because there are sequences that are very unusual in stems, but common in inflected forms. For instance, [ts] is rare in stems (e.g. *Katz*, *Hertz*) but is ordinary in inflected forms like *cats* or *hurts*. I argue below that Seuss indeed uses [ts] as a basis for coinages.

All the analytic work I did for this article (lexical databases with phonetic transcription, R scripts, spreadsheets) may be accessed in the Supplemental Materials for this paper, currently at <https://linguistics.ucla.edu/people/hayes/Seuss/SupplementalMaterials.zip>.

taxonomy (Fordyce 1988:238-239), I am largely addressing phonesthemes, whereas Shih et al. and Kawahara are addressing sound symbolism.

2.1 What principles might be used to characterize the coined words?

Following a few pilot efforts, I settled on the following procedure to guide the work: I searched a fairly large preliminary set of predictive principles, then narrowed it down to a smaller set with just the most effective ones.

My pilot studies indicated that several strongly-predictive traits consisted of *word-initial syllable onsets*, such as the [sn-] of *Snumm*. To be thorough, I searched the entire set of 73 attested word-initial onsets, irrespective of whether they occur in the real-word corpus or the Seuss corpus (some onsets occurred only in Seuss). I also found that particular vowels, occurring in main-stressed position of a word, were sometimes highly predictive of Seuss, such as [ʌ]. Thus, in my more serious search, I included all 15 main-stressed vowels from the corpora as potential predictive factors.²

I also incorporated into my search some ideas taken from the research literature on sound symbolism, mostly from work on **phonesthemes**. These are short segmental sequences that don't fully qualify as morphemes, but nonetheless often impart a (perhaps vague) meaning to words that they contain. Phonesthemes are discussed extensively below in §3.4. For purposes of including multiple phonesthemes in the initial search, I relied on the lists in Marchand (1969) and Hutchins (1998).

2.2 The culling procedure

The final model was made by reducing the original set of factors, described in the previous section, to a smaller set, each of whose members demonstrably contributes to predicting Seussian status in a logistic regression model.

I offer a brief note on logistic regression. Every principle that might help predict Seussian status is here termed a **feature**.³ Each feature is given a particular number, called its **weight**. In my own setup, if the weight of a feature is positive, it means that the feature favors Seuss-coinage status; if the weight is negative, it means that the feature militates against Seussian status; and if it is zero, the feature is indifferent. Greater magnitudes of weights (either positive or negative) have greater effect.

The output of the model, for any given phonetically-transcribed word, is a value ranging from zero to one, expressing the estimated probability that a word is Seussian; a perfect model would assign one to all Seuss coinages and zero to all ordinary words. Where computation enters

² For both cases above, it might have been possible to *generalize* the initial findings, using standard phonological features in the normal way. However, I usually found this was not all that helpful (see fn. 7 for the one case that I retained in the final model). Since generalizing the features vastly expands the number of hypotheses to consider and makes diagnosis harder, I did not pursue it any further.

³ The terminology comes, I believe, from computer science. In linguistics, we would be more likely to call the features **constraints**, following the research tradition of Optimality Theory (Prince and Smolensky 1993/2004). However, constraints normally act only as penalties for particular candidates, and here they more often reward them; hence the terminology.

is in setting the weights: one's chosen logistic-regression software will calculate the weights that best separate out the Seuss coinages in the data from the real words.⁴

My chosen software was the `bayesglm()` function within the R statistics system (<https://cran.r-project.org/web/packages/arm/arm.pdf>). This version of logistic regression is somewhat conservative, assigning lower-magnitude weights than would obtain under the simplest forms of logistic regression.

I sought to trim my large candidate system of features into one that would be much smaller but perform almost as well. First, I removed all constraints that tested nonsignificant (by a .1 *p* value), then culled further with the `stepAIC()` function, which lets us keep a feature only if it creates improvement by the Akaike Information Criterion, a well-known measure that penalizes overly complex models.⁵ Statistical evaluation of all models reported here is given below in Appendix A.

2.3 The features of the completed model

The tables in (3) and (4) below give what I found. There is one non-specific feature in the system, the **Intercept**, which simply is a raw penalty against being Seussian — a sensible penalty, in light of the disparity in numbers. The weight of the Intercept is -4.80 , which is quite large. Hence, for any form to receive a really strong Seussian probability, it must rack up substantial compensation from the positive-weighted features in order, as it were, to climb out of the hole.

For each feature in (3) and (4), I give the following information:

- The **form** of the feature; usually a sequence of phonemes. “[” means that a feature counts the relevant sequence only if it is initial in the word; “]” analogously means “final”; no bracket means “anywhere”. A few features deviate from this format and are described in words.
- The **weight** of the feature. For intuitive interpretation of weights see footnote.⁶
- The **number of words**, both Seussian and real, that come under the scope of the feature, with representative Seussian examples.

⁴ The method of computation used here is a standard one, namely to maximize **likelihood**, the probability predicted by the model for the corpus as a whole. This is the product of the probability assigned to Seussian-status for all 435 Seuss coinages, multiplied by the probability assigned to real-status for all 17,744 real words. Maximizing this product (it would approach 1 in a perfect model, 0 in a perfectly-bad model) maximizes what we intuit as model effectiveness. Actual likelihood achieved by this and other models is reported (in log form) in Appendix A.

⁵ It would have been more principled to use `stepAIC()` to do *all* of the culling, but this task overwhelmed my computing equipment.

⁶ By the math of logistic regression, it can be shown that the formula $e^{-weight}$ tells us the *ratio* of predicted probability between otherwise-identical forms that do and do not come under the scope of a feature. A weight difference of 5 means a probability ratio of about 150, 2 about 7, 1 about 2.7, .5 about 1.6, and 0 means equal probability.

- **Explanatory comments**, where applicable; these serve as placeholders for the discussion to follow. “Marchand” with page number means that a sequence has been identified as an English phonestheme by Marchand (1969), the reference source used for statistical testing in §4.

(3) *Table: features that favor being a Seuss coinage*

	Feature	Wght.	Seuss	Real	Examples	Comment
a.	pf	6.95	3	0	<i>Dumpfer, Humpf, Klopfer</i>	German
b.	[ʃ]	6.88	5	0	<i>schlopp, Schlottz, Schloppity</i>	German
c.	[θn]	6.65	3	0	<i>Thnad, Thnadner, Thneed</i>	phonotactically illegal
d.	[ʃn]	5.84	2	0	<i>Schnack, Schnutz-berry</i>	German
e.	[v]	5.17	1	0	<i>(Van) Vleck</i>	phonotactically illegal
f.	[zl]	5.17	1	0	<i>Zlock</i>	phonotactically illegal
g.	[vr]	4.38	2	0	<i>Vroo, Vroom</i>	phonotactically illegal
h.	[θw]	4.12	1	1	<i>Thwerll</i>	phonotactically marginal
i.	[z]	3.31	40	48	<i>Zax, Zans, Zuff</i>	Marchand 333
j.	[sn]	3.20	21	41	<i>snop, snarggle, Snarp</i>	Marchand 331-332
k.	2 medial stressless syls	3.17	19	282	<i>Bippo-no-Bungus, Jill-ikka-Jast, Fotta-fa-Zee</i>	metrical
l.	[gw]	2.56	1	4	<i>Gwark</i>	phonotactically marginal
m.	ts	2.50	12	12	<i>Gitz, Glotz, Schlottz</i>	German
n.	ap]	2.44	9	21	<i>Dopp, G-r-r-zopp, Yopp</i>	Marchand 314
o.	ts]	2.24	10	2	<i>Dill-ma-dilts, Gitz, Skritz</i>	German
p.	əmə	1.96	8	90	<i>filla-ma-dill, Fizza-ma-Wizza-ma-Dill, Katta-ma-side</i>	metrical
q.	ks]	1.80	12	88	<i>Bix, Gox, Grox</i>	
r.	ps]	1.74	2	19	<i>Bopps, plupz</i>	
s.	[kw]	1.70	7	89	<i>Quan, Kwong, quirkle</i>	Marchand 330
t.	ʌ	1.68	99	973	<i>Dutter, Flunn, Umbus</i>	Marchand 316
u.	[gl]	1.66	12	59	<i>Glotz, GLUGG, glump</i>	Marchand 327
v.	u:	1.55	59	888	<i>Foodle, Katroo, Oobleck</i>	Marchand 316
w.	[j]	1.51	15	149	<i>Yook, Yekko, Yertle</i>	Marchand 327
x.	ʌp	1.39	9	32	<i>Flupp, ga-fluppt(ed), Glupp</i>	Marchand 314
y.	[bl]	1.37	12	71	<i>Blogg, blunk, blurp</i>	Marchand 323
z.	[fl]	1.27	10	122	<i>Flobber, Flupp, flubbulous</i>	Marchand 328
aa.	[skr]	1.25	5	33	<i>Skrux, Skrink, Skritz</i>	Marchand 326
bb.	ɪŋk	1.21	10	55	<i>Blinkey, Dinker(ville), Dinkzott</i>	Marchand 336
cc.	[n]	1.20	21	370	<i>Nobsk, nop, Natch</i>	
dd.	[w]	1.20	19	277	<i>Wocket, Woset, Wah-Hoo</i>	Marchand 329
ee.	p]	1.06	36	280	<i>Bip, blurp, bloop</i>	Marchand 314
ff.	f]	1.02	14	136	<i>Olf, Dawf, Feffer-feffer-feff</i>	

gg.	[f	0.86	20	541	<i>Fonz, Fotichee, Fotta-fa-Zee</i>	Marchand 328
hh.	[g	0.81	23	257	<i>Gox, Gargel-orum, Gack</i>	Marchand 327
ii.	ɪ	0.81	88	2494	<i>Biffer-Baum, Slippard, Ziff</i>	Marchand 316
jj.	[gr	0.78	10	135	<i>Grox, Gractus, grum</i>	Marchand 327
kk.	Voiced obstruent	0.58	298	9390	<i>Obsk, Ubb, Umbus</i>	see fn. ⁷
ll.	əɪ]	0.46	38	1442	<i>Biggel, Bingle, Yertle</i>	Marchand 316

(4) *Table: features that favor being a real word*

	Feature	Wght.	Seuss	Real	Examples	Comment
a.	Syls > 2	-0.82	75	8122	<i>Katta-ma-side, Eisenbart, Boliver</i>	
b.	V-initial	-1.05	14	3533	<i>Obsk, Antrum, Ubb</i>	
c.	[s	-1.11	7	1083	<i>Sard, Solla, Sollew</i>	
d.	[r	-1.47	4	1070	<i>Ronk, Redd-Zoff, rippulous</i>	Marchand 332
e.	[k	-1.74	3	1357	<i>Katta-ma-side, Katroo, Keck</i>	Marchand 325
f.	[ʃ	-2.19	0	146	none	Marchand 333
g.	[st	-2.55	0	198	none	Marchand 324

2.4 How well does the model work?

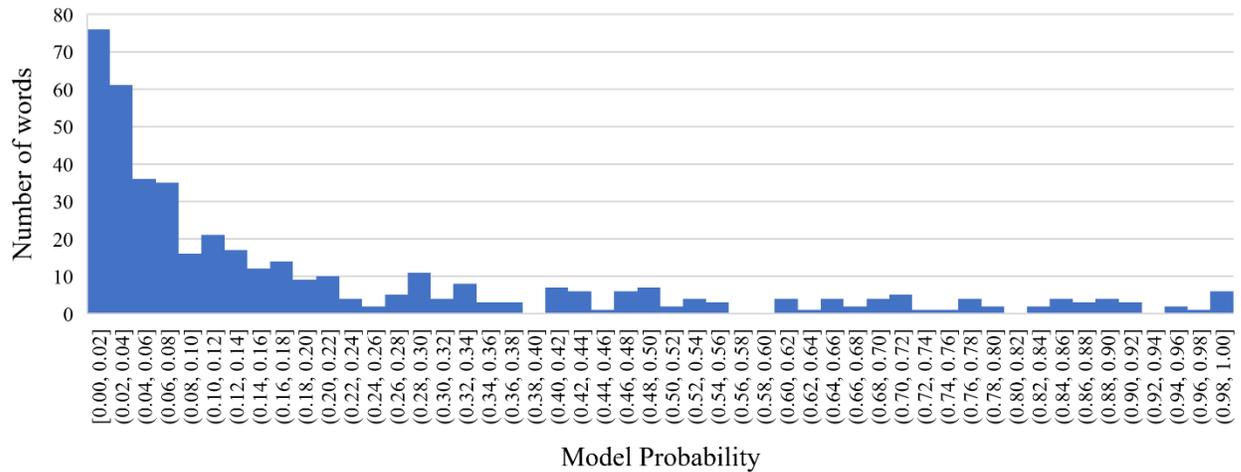
To be clear on this point, the model cannot reliably distinguish Seuss coinages from normal words. Most particularly, it is unable to achieve anything like a perfect score (i.e., 1) for most of the Seussian coinages. Nevertheless, I believe it makes useful distinctions. The model's average "probability is Seuss" for Seussian coinages is 21.1%, whereas the average "probability is Seuss" for normal words is only 1.9%.

We can get a more detailed picture by comparing histograms. In (5), I plot the probabilities assigned by the model to the 435 Seuss words, compared to the probabilities assigned to the 17,744 real words (in the latter, the scale is compressed to accommodate them in the same vertical space).

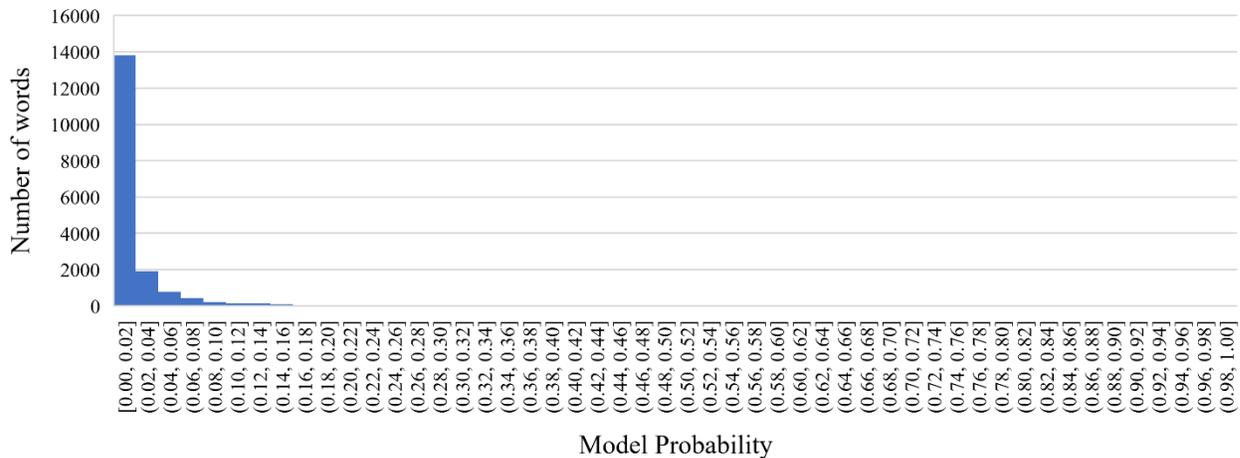
⁷ The weight is low, but this is because there are nine other features in the system that do the same work, essentially preempting this general constraint.

(5) Histograms: comparing probabilities assigned to the Seuss coinages and to real words

a. Seuss coinages



b. Real words



It is clear that while the model falls far short of perfect (this would be: sharp spikes at the right for Seuss and at left for real), it does predict a very different distribution for the Seussian coinages.

We can also examine the extremes of behavior. In (6) are given the ten most “Seussian” Seuss coinages. I include the particular phoneme sequences that are picked up by the features of (3) and (4) and converted, via the math of logistic regression, to high predicted probability.

(6) The 10 *Seuss* coinages with the highest model probability

Coinage	Relevant phonological properties	Model score
<i>Schnutz</i>	[[fn], [ts], [ʌ]]	0.999
<i>Schlottz</i>	[[ʃ], [ts]]	0.999
<i>schlupp</i>	[[ʃ], [ʌp]]	0.998
<i>schlopp</i>	[[ʃ], [ʌp]]	0.996
<i>Humpf</i>	[ʌ], [pf]	0.992
<i>shlump</i>	[[ʃ], [ʌ], final [p]]	0.992
<i>Zatz</i>	[z], [t], voiced obstruent	0.979
<i>Zomba-ma-tant</i>	[z], [əmə], medial lapse, voiced obstruent, syls over 2	0.959
<i>Yuzz-a-ma-Tuzz</i>	[j], [ʌ], [əmə], medial lapse, voiced obs., syls over 2	0.954
<i>Thnad</i>	[θn], voiced obstruent	0.918

In less detail, (7) gives the ten *least* Seussian Seuss coinages, as well as the “most Seussian” and “least Seussian” real words (the latter consists of ten words randomly chosen from the 484 Seussian words that got a 0.000 score).

(7) *Further examples of the model’s behavior*a. *Seuss coinages rated by the model as minimally Seussian*

lopulous (0.004), *Lass-a-lack* (0.004), *Hippo-Heimer* (0.004), *Ronk* (0.003), *Antrum* (0.003), *Offt* (0.003), *Gee-Hossa-Flat* (0.003), *Solla (Sollew)* (0.003), *rippulous* (0.002), *Keck* (0.001)

b. *Real words rated as most Seussian by the model*

xerox (0.845), *zinc* (0.843), *quartz* (0.837), *zigzag* (0.837), *waltz* (0.757), *snuggle* (0.754), *snuff* (0.749), *snoop* (0.732), *zip* (0.722), *snub* (0.659)

c. *Sample of real words rated as minimally Seussian by the model (all at 0.000)*

*administration, appreciation, appreciative, chicanery, competitor, elaboration, electromagnetic, encyclopedic, immemorial, meteorological*⁸

These are meant mainly as a guide for the intuition, though the forms of (7b) evoke a further phenomenon: Seuss occasionally adapts a real word, often bearing Seussian phonological traits, to serve as a novel word; for example *zip*, respelled as *Zipp*, is used as a surname in *Oh Say Can You Say?*. These forms are discussed in Appendix B.

One wonders whether the model could be improved by further work. I would judge that this is likely, since many of the Seussian coinages are assigned low scores but somehow sound

⁸ Perhaps more informative is a sampling of minimal-Seuss-score monosyllables, which are for the most part not from learned vocabulary: *case, coal, cork, corn, course, shake, stake, stave, stove, sty*.

Seussian to me, for example *sporn*, *Jounce*, *tweetle*, all with scores below 0.02 — something is still missing. However, I believe the model in its present form suffices for its intended purpose; namely, that we can inspect it, trying to find in its features some principles that will be informative about Seuss's coinage practice in general terms.

3. The Seuss coinages: seeking general principles

I will put forth four proposed principles of Seussian coinage.

3.1 Meter

First, the Seussian words are skewed somewhat to make them fit easily into Seuss's favorite meter, anapestic tetrameter — see features (3k) and (3p) above. Since these metrical principles are so distinct from the main theme of this paper, I have relegated discussion of them to Appendix C below.

3.2 Phonotactic violations

As Nilsen (1977) observed, a noticeable minority of the coinages violate principles of English phonotactics; specifically word-level phonological well-formedness. For English phonotactics see e.g. Hammond (1999), Hayes and Wilson (2008), and Daland et al. (2011). The patterns noted below are probably not controversial.

First, a number of onsets found in the Seuss coinages are not permissible in the core English vocabulary (although they may occur in unassimilated borrowings).

(8) Illegal onsets occurring in Seuss coinages

<i>Onset</i>	<i>Count</i>	<i>Examples</i>
[ʃl]	5	<i>schlopp</i> , <i>Schlottz</i> , <i>schlupp</i>
[θn]	3	<i>Thnad</i> , <i>Thnadner</i> , <i>Thneed</i>
[ʃn]	2	<i>Schnack</i> , <i>Schnutz</i>
[vl]	1	<i>van Vleck</i>
[zl]	1	<i>Zlock</i>
[vr]	2	<i>Vroo</i> , <i>Vroom</i>

Among other onsets tagged in the feature-selection process described above, [θw] and [gw] are very unusual in real English words and might be regarded as hovering on the fringes of ill-formedness. The impossible coda [bsk] is found in *Obsk* (a bird in *Scrambled Eggs Super*) along with the coined words of the same passage that rhyme with it.

The letter name *Nuh*, from *On Beyond Zebra*, ends in a lax vowel ([nʌ]),⁹ something impossible in ordinary words and limited to quasi-gestural forms like *uh* ([ʌ]; hesitation noise) and *duh* ([dʌ]; used to indicate one's interlocutor has missed something obvious).

⁹ We know this because in *On Beyond Zebra* *Nuh* is the letter used to spell *Nutches*, which rhyme with *hutches*.

Lastly, consider *Snumm*, quoted above in (2). Along with *Snimm* (a proper name from *Too Many Daves*) this coinage violates a phonotactic principle discussed in Davis (1991): English avoids the occurrence of similar or identical consonants in the C positions of the formula **sCVC**. Davis's constraints include, for instance, bans on /spVp/ and /skVk/ (*spip* or *skeck* would be odd as English words). In the present context, the relevant ban, also noticed by Davis, is on /sNVN/, where N is any nasal consonant. As Davis points out, no such words exist in English and I personally find *smem*, *smun*, *snam* (and indeed *Snumm* and *Snimm*) to sound odd.

Unsurprisingly, none of these phonotactic violations are extreme, like, say, the use of uvular consonants or grossly sonority-violating initial or final clusters. It seems that Seuss wanted his words to sound funny, but would hardly want to inflict an impossible phonetic challenge on his readers.

3.3 Word that sound German

Nilsen (1977) and Teuber (2018) suggest that a number of the Seuss coinages sound like German words. Some of these have already been mentioned in the previous section: words beginning in [ʃ] and [ʃn] are aberrant in English, but are normal in German.

German is of course closely related to English and has similar phonotactics. Yet the phonological history of the language (see e.g. Chambers and Wilkie 1970) has produced some points of departure. By the Second Consonant Shift, historical Germanic *t and *p (preserved intact in English) evolved in certain contexts into [ts] and [pf], sequences that are very rare in English. Thus, the clusters in *Katze* ['katsə] 'cat' and *Kropf* '(bird's) crop' are a point of phonological divergence between German and English that is attested in multiple words. The [ʃ] and [ʃn] clusters just mentioned were also created by sound change, from historical *sl and *sn. All four of these German-linked patterns are shown in (3) above to be statistically unambiguous features for Seuss coinages.¹⁰

That these coinages were actually intended by Seuss to sound German is made plausible by several factors. First, the orthography Seuss chose for them is largely German, as in *Gitz*, *Glutz*, *Schlottz*, *Schnutz* (that is, *tz* not *ts*, *sch* not *sh*). Second, the texts include a few overt German cultural references, notably the blue-footed mandolinist *Gretchen von Schwinn*, from *Oh Say Can You Say*, and the *castle of Krupp*, from *Dr. Seuss's Sleep Book*. Lastly, Seuss's German-styled coinage practice can be related to his own life history (Morgan and Morgan 1995): he grew up in a German-speaking family (he was third-generation) in Springfield, Massachusetts, a city that during his youth included a vibrant German-American community.

3.3.1 The German coinages and the American audience

It is only natural that Seuss, a popular artist, would have attempted to create coinages that would make sense to his readers. In the present context this raises the question of whether Seuss's audience (mostly mid-century Americans) would have been able to identify Germanness

¹⁰ I did not include in my feature set any phoneme sequences that correspond to actual German morphemes, but these do occur a number of times in the coinages: *Herk-heimer*, *Hippo-heimer*, *Eiffelberg*, *Eisenbart*, *Bickelbaum*, *Biffer-baum*, *Katzen-bein*, *Katzen-stein*, *Spritz*.

in nonce words. An intriguing research finding by Oh et al. (2020) bears on this question: they show by experiment that non-Māori residents of New Zealand, very few of whom can actually speak Māori, nonetheless have an accurate sense of the phonotactic principles of the language, obtained from second-hand exposure. This suggests that if Seuss's audience had enough second-hand exposure to German they likewise could have internalized a sense of what German phonology is like. It seems reasonable to me to claim that mid-century Americans did indeed have considerable exposure to German; this was the period following World War Two, and closer to the historical time when German-Americans were the nation's largest ethnic minority.¹¹ Of course, even now many American Seuss readers would surely recognize *Schlottz* as a German-like word.¹²

3.4 Phonesthemes

For present purposes, I define a phonestheme as the following: (1) it is a segment or segment sequence that occurs in multiple words; (2) it has some vague, often expressive meaning; (3) its "residue" in a word is not a morpheme; e.g. in word of the form [*Ph X*]_{word}, where *Ph* is a phonestheme, *X* is not in general an identifiable morpheme of the language. To give an example, initial [sn-] is a well-known phonestheme of English. Its meaning is (vaguely, as always) 'having something to do with the nose', as in *snoot*, *snot*, *sneeze*, *snout*, *snuff*, *snore*, *sniff*, *sniffle*, *snort*; and by extension 'looking down the nose at', *snob*, *snooty*, *sneer*, *snicker*, *snide*, *sniffy*, and *snub*.¹³ We will examine other phonesthemes below.

Phonesthemes are the topic of an enormous research literature,¹⁴ which I briefly discuss before going on to the Seuss coinages.

Some phonesthemes are said to have a natural phonetic basis, as in the affiliation of [i] (a low-sonority vowel with high F2) with smallness. For a careful overview of this topic see Kawahara (2021b). For present purposes I believe it will be safe to ignore whether a phonestheme is natural or arbitrary.

An element of phonesthemes that will be essential here, however, is the question of **style**. I adopt what appears to be a minority position, that of Bloomfield (1933:156) and Joseph (1994),

¹¹ For what it's worth, I myself was a mid-century American and judge that during my childhood the United States was far more attuned to Germanness than it is today. The mid-20th century was the heyday of Wernher von Braun, "Hogan's Heroes," beers like *Schlitz*, *Pabst*, *Schmidt*, and *Rheingold*, stores with names like *Ski Haus* and *Cheese Haus*, and a fast-food chain called *Wienerschnitzel*.

¹² A final note on Germanness: we will never know what the tight-lipped Seuss thought about his Germanisms, but they could not have been a fully neutral topic for him. In 1917, when the United States joined the fight against Germany, a xenophobic campaign to deculturalize German-Americans took place, including in Springfield, where the young Seuss was beaten in the street (Morgan and Morgan 1995).

¹³ Throughout, I limit examples to words I know, hopefully thereby approximating the words Seuss knew. The *Oxford English Dictionary* offers a wealth of obscure and obsolete words that include one of the three focus phonesthemes discussed below ([sn-], [z-], [j-]). I believe including these words would lead to similar conclusions.

¹⁴ Some helpful entry points to this literature are Schmidtke et al. 2014 (psychology and cognitive science), Magnus 2001 (linguistics), and Kawahara 2020 (linguistics). On the role of phonesthemes in word coinage (i.e. by ordinary people of the past) see Pentangelo (2020). The idea that Seuss uses phonesthemes in his coinages was first put forth by Teuber (2018), who, however, does not assess it.

who emphasize the stylistic function of phonesthemes: phonesthemic words are characteristically vernacular in tone and expressive in function. Joseph (pp. 222, 229) articulates this view clearly, describing phonesthemic words as “expressive, affective, connotative”; they “add color to the language.”

Style is especially important when we consider word coinage. Speakers obviously do not concoct phonologically-novel words for the purpose of making their meaning clear; rather, these coinages are intended to make an impression, based their imaginative phonological content. Earlier (§1), I mentioned the apparent fact that many of our existing words originated as phonesthetic coinages, the work of creative speakers long forgotten. It is not unreasonable to regard these coinages, at least at the moment of origin, as the deployment of phonesthemes in the creation of verbal folk art. I suggest that Seuss embraced this art form in forming his own distinctive and highly vernacular style.¹⁵

3.4.1 *Phonesthemes and style: a three-way distinction*

Before plunging into the Seuss data, I will offer a hypothesis about phonesthemes which seems to hold good for his practice — and is perhaps true of phonesthemes in general. The idea is that for any phonestheme, we will normally find words that fit into each of the following categories.

(9) *A three-way classification for phonesthemic vocabulary*

- a. Words in the **meaningful core** of a phonestheme (“core words”) both contain the phonestheme and bear the appropriate meaning.
- b. Words in the **penumbra** of a phonestheme contain the phonestheme and also convey the vivid, expressive character of phonesthetic style; but they do not bear the meaning of the phonestheme.
- c. Words in the **neutral zone** of a phonestheme contain the segments of the phonestheme but do not bear the meaning of the phonestheme and lack a vivid, expressive meaning; they are not phonesthetic.¹⁶

¹⁵ The idea that phonesthemes are devices of folk art is not the only possible approach. An alternative is the “word affinities” approach, put forth by Bolinger (1965) and Magnus (2001). This sees phonesthemes as the result of word-comparison: human language learners comb through their lexicons, seeking all conceivable correlations between phoneme sequences and meaning (this scheme that has since been computationally modeled; Otis and Sagi 2008, Liu et al. 2018). When pursued to a successful conclusion, the search yields knowledge of the authentic morphology; phonesthemes are the morpheme candidates left on the workbench when learning doesn’t fully succeed. The word-affinities approach strikes me as plausible, but incomplete, for it misses the expressive function. A purely distributional, expressionless phonestheme, if such exists, would have no value for a novel coinage and would not be employed as such. Moreover, expressiveness alone suffices for using a phonestheme even when meaning is absent; see discussion of the “penumbra” below.

¹⁶ Citations: (9a) is agreed upon by all. To my knowledge, only Joseph (1994:229-230) has ever noticed (9b), penumbra. The neutral zone, (9c), is widely noted, e.g. by Jespersen (1922:406) and Fordyce (1988:177). Note that my use of the term “core” differs from that of Fordyce, who uses it to describe those words that embody the

I illustrate this taxonomy for the “nasal” phonestheme [sn-] already mentioned. The core of this phonestheme would include the words with nasal meaning enumerated earlier: *snoot*, *snot*, *sneeze*, *snout*, *snuff*, *snore*, *sniff*, *sniffle*, *snort*, *snob*, *snooty*, *sneer*, *snicker*, *snide*, *sniffy*, *snub*. What of the penumbra? I suggest that it includes words like *snazzy*, *snag*, *snatch*, *sneak*, *snip*, *snitch*, *snoop*, and *snug*. These seem unnasal in their meaning, but they are nonetheless expressive, in the way that phonesthemes characteristically are. To defend this claim, I juxtapose some words that occupy the penumbra of the [sn-] phonestheme with their literal near-equivalents:

(10) *Comparing penumbral phonestheme words with literal expression*

<i>snazzy</i>	vs.	<i>stylish</i>
<i>snatch</i>	vs.	<i>abruptly grasp</i>
<i>snip</i>	vs.	<i>cut deftly</i>
<i>snoop</i>	vs.	<i>spy</i>
<i>snug</i>	vs.	<i>comfortable</i>

Here are further comparisons, in this case involving words in the penumbrae of two phonesthemes to be discussed below, [z-] and [j-]:

(11) *Further comparisons with [z-] and [j-]*

<i>zit</i>	vs.	<i>skin blemish</i>
<i>zilch</i>	vs.	<i>nothing</i>
<i>yank</i>	vs.	<i>pull sharply</i>
<i>Yankee</i>	vs.	<i>American</i>

I think it is clear that the words in (10) and (11) that include the phonestheme are more vivid and more colloquial. The implication is that a phonestheme does *not* require its core meaning to be present to render its stylistic effect.

Unsurprisingly, the element of vivid style that is the sole phonesthemic property of penumbral words is also found in the words of the core, as the examples of (12) suggest.

(12) *The stylistic effect of phonesthemes in core vocabulary*

<i>snoot</i>	vs.	<i>nose</i>
<i>snot</i>	vs.	<i>mucus</i>
<i>snub</i>	vs.	<i>disdain</i>
<i>zany</i>	vs.	<i>highly eccentric</i>
<i>zap</i>	vs.	<i>strike very suddenly</i>
<i>zip</i>	vs.	<i>proceed briskly</i>

phonestheme’s meaning most clearly and saliently. That degree of adherence to the meaning of a phonestheme is gradient seems clear from Fordyce’s as well as Hutchins’s (1998) experiments.

yell vs. *shout*
yahoo vs. *hooray*
yammer vs. *complain*

Consider next the neutral zone. It is treated here as the set of words that accidentally contain the segments of a phonestheme, in the same way that, say, *lens* accidentally contains the [-z] of the plural suffix. This zone can be a source of frustration to anyone lecturing about phonesthemes, whose audience is naturally inclined to ask, “What about word X? Isn’t that a counterexample?” It seems best to acknowledge that most phonesthemes do have a neutral zone, but the existence of this zone should not be taken as counterevidence to the existence of the phonestheme — in pointing out a phonestheme, we are only pointing out a pattern that is too frequent to be coincidence, not an implicational law. Indeed, Hutchins’s (1998) experiments affirmed psychological reality for phonesthemes that possess a demonstrable neutral zone. As for the neutral zone of [sn-], I suggest that two plausible candidate words are *snow* and *snail*.

A number of scholars (e.g. Jespersen 1922:407, Malkiel 1990, Magnus 2001:8, 72, Pentangelo 2020)) have suggested that phonesthemes exert a kind of gravitational attraction; drawing additional words into their membership by adjusting either their form¹⁷ or their meaning. In present terms this claim can be elaborated a bit: I suggest that members of the periphery may gradually assume semantic properties of the core, and members of the neutral zone may be drawn into the core or periphery, becoming regarded as phonesthetic and expressive. Such drift is likely the result of language misacquisition; children are prone to mislearn either the style level or the meaning of phonesthetic words.

Here is one example of drift into the core: Malkiel (1990:99-110) documents an Italian phonestheme of the form CVC_iC_iV (C_iC_i a geminate) with meaning ‘negative, or ridiculous, or both’, which has pulled in words that were formerly neutral such as *nullo* ‘nothing’ and *secco* ‘dry’, giving them novel secondary usages that fit the core meaning. For drift into the penumbra, I am on much more speculative ground, but the reader may wish to ponder the words *snooker* and *snipe*. I feel that they belong in the penumbra, not the neutral zone, of [sn-]: as words they seem absurdly jokey and vivid for the purpose of denoting an ordinary indoor sport and bird species.¹⁸ The Broadway composer Irving Berlin evidently felt a sense of pull for the phonestheme [j-] when he wrote the musical *Yip Yip Yaphank*, attracting the name of the Long Island town where he did his Army service ([ˈjæpæŋk]) into the penumbra of the [j-] phonestheme.

What enables a neutral-zone word to resist the inward pull of its component phonestheme? I suspect frequency matters: in my lexical database, the most frequent words (per CELEX) beginning with the phonesthemes discussed here have at most a modest penumbral tinge: *snow*,¹⁹ *Z*, *zone*, *year*, *use*, *young*. The other cause of phonestheme-resistance is speech register: formal

¹⁷ For example, Jespersen suggests that *peep* originated as a phonesthetic ‘repair’ of *pipe*, a word which had lost its phonesthetic appropriateness when the Great Vowel Shift altered its vowel from [i:] to [aɪ].

¹⁸ It is circular reasoning, but worth pointing out, that Seuss used both words in his books: *Snipe* appear (as such) in *If I Ran the Circus*, and *Snookers* is a surname in *Happy Birthday to You*.

¹⁹ *Snow* is perhaps a core word for sniffers of white powder cocaine, a usage dated to 1914 by the *OED*.

or technical words are incompatible with the stylistic character of phonesthesia, and so they can contain the phonesthemic sequence without being pulled in: *Snell's Law*, *zoning*, *zinc*, *zinnia*, *yarrow*, *ubiquity*. For further discussion see Magnus (2001:5, 10, 34, 72).

An important implication of the above for present purposes is this: *a word coined for purposes of writing a children's book would be unlikely to occupy the neutral zone of any phonestheme it contains*. If an author uses the segments of phonestheme, it will probably be perceived by readers as being intended as a phonestheme. The word frequency of a coinage is very low (i.e., zero); and technical or formal vocabulary would hardly be expected in a children's book.

To sum up this section, I will approach Seuss's coinages from the viewpoint of the three-way taxonomy of (9), which emphasizes (a) the stylistic role of phonesthemes; (b) the possibility of phonesthemes that convey style but not the relevant meaning; (c) gravitational attraction, under specified conditions, from neutral zone to penumbra to core.

We turn now to the some particular phonesthemes that I claim are used in Seuss's coinages. I will argue that a minority of the phonesthemic usages in Seuss are core, the rest are penumbral, and none are neutral.

3.4.2 [sn-]

The 21 Seuss coinages that begin with the "nasal" phonestheme [sn-] are given in (13):

(13) Coinages with [sn-]

snop, snarggled, Snarp, snaff, Snux, Snumm, snuv, Snegg, Sneth, Snick, Snimm, Snee, Sneetch, Sneetcher, Sneedle, Sneeden, Sneelock, Sneepy, Snooker(s), Snoo, Snoor

Of these, I have identified four as belonging to the core of [sn-]. *Snaff* is an obvious case, as it describes an acting of vigorous sniffing by the boastful bear in *The Big Brag*. *Snarggled* appears in a sequence of verbs with *sneezed*, *snuffled*, and *sniffed*, describing inhalation of polluted air, in *The Lorax*. The snobbish *Sneetches* plainly qualify, per Seuss's description:

(14) With their snoots in the air, they would sniff and they'd snort

A more subtle case is the *Sneedle*, from *On Beyond Zebra*: this is an insect whose nose takes the form of a large and frightening stinger:

(15) Then we go on to SNEE. And the SNEE is for Sneedle A terrible kind of ferocious mos-keedle Whose hum-dinger stinger is sharp as a needle.

However, this seems to exhaust the core [sn-] words in Seuss, as the remaining 18 [sn-] coinages are slim pickings for anyone seeking out nasal meaning. For instance, the *Drum-Tummied Snumm*, from (2) above, has a spectacular tummy, but a very ordinary nose. Elsewhere in *If I Ran the Circus*, neither the *Harp-Twanging Snarp* nor *Mr. Sneelock* seem nasal in any

way, and the same goes for the remaining words in (13). I would suggest that these forms are indeed penumbral; i.e. expressive but not nasally meaningful.

3.4.3 [z-]

[z-] is given short shrift by my primary reference, Marchand (1969) (“an infrequent initial”) but is taken more seriously by Wescott (1980), who demonstrates considerable productivity for it. Let us consider the cases from my own data corpus. The 47 real [z]-initial words in my dictionary include eight that seem fairly clearly phonesthetic: *zest*, *zigzag*, *zing*, *zip*, *zoom*, *zot*, *zany*, and *zap*. Were I try to define the core meaning of [z-], I would guess something like “with great liveliness”. Thus, a person who is *zany* is not just somewhat crazy, but crazy in a lively way; for a lizard to *zap* an insect it must make a very abrupt movement of its tongue.

The [z-] phonestheme also appears to have a penumbra. For example, *zit* is a very expressive way to denote a pimple, but pimples are not lively. *Zilch* means ‘nothing’, but is used to express the idea with feeling and humor. *Zonked* is plainly expressive but denotes stupor rather than liveliness. There is a neutral zone, composed of technical expressions like *zinc* and *zinnia*. A possible example of a neutral-zone word drawn toward the penumbra is *Zenith*, a brand name that did well for selling television sets in Seuss’s day.

[z-] has been noticed before by scholars of the Seuss coinages (Teuber 2018, Keys 2021) and is indeed the most frequent phonestheme in his work, with 40 occurrences.

(16) Coinages with [z-]

Zomba-ma-tant, *Zozzfozzel*, *Zax*, *Zans*, *zang*, *Zatz*, *Zatz-it*, *zazz*, *Zuff*, *Zuk*, *zum*, *zummer*, *Zummz*, *Zummzian*, *Zutt*, *zuzz*, *Zall*, *Zong*, *Zorn*, *Zower*, *Zike*, *Zed*, *Zellar*, *Zelf*, *Zable*, *Zayt*, *Zidd*, *Ziff*, *Zillow*, *Zinn-a-Zu*, *Zind*, *Zinzibar*, *zizz*, *Zizzer-Zoof*, *Zizzy*, *Zizzer-Zazzler-Zuzz*, *Zeep*, *Zook*, *Zooie*, *zoop*

Searching through their meanings, we again find just a few cases in which the [z]-word occupies the phonesthemic core. *Zoop* is part of *Zoom-a-zoop*, describing a spectacularly virtuosic trapeze act in *If I Ran the Circus*. When the bird character Gertrude McFuzz suddenly sprouts a spectacular tail to satisfy her vanity, she does it “With a *zang*! With a *zing*!”. With an extension to not-quite-initial position, we may include *G-r-r-zapp*, *G-r-r-zibb*, *G-r-r-zopp*, the sounds of the arrows shot by the Yeoman of the Bowman in *The 500 Hats of Bartholemew Cubbins*. However, most instances of [z-] in Seuss appear to be only penumbral. Notably, several [z]-initial Seussian animals are placid and serene: the *Ziffs* and *Zuffs* of *Scrambled Eggs Super*, the *Zizzer-Zazzler-Zuzz* of *Dr. Seuss’s ABC*, the *Zatz-It* of *On Beyond Zebra*, and the *Zans* and the *Zeep* of *One Fish Two Fish Red Fish Blue Fish*.²⁰

²⁰ A possibility to consider is that the [z-] phonestheme possesses two cores, the second of which evokes sleepiness. Some relevant real words I have noticed are *zzz* (orthographic phonestheme denoting snoring), *zone out*, and *zonked*. For Seuss, several of the animals cited above are portrayed as sleepy. For multiple-core phonesthemes see Fordyce (1988:194-195), Magnus (2001), and Pentangelo (2020).

3.4.4 [j-]

For Marchand (1969:334) this phonestheme is for ‘words expressive of vocal sounds’; my own preference would be to characterize it as ‘vigorous or uncontrolled vocalization’. Core examples are: *yahoo*, *yammer*, *yatter*, *yap*, *yawp*, *yell*, *yelp*, *yip*, *yipe*, *yippee*, *yo*, *yowl*, and *yodel*. Some penumbral words are *yo-yo*, *yank*, and *Yankee*. [j-] is a “weaker” phonestheme than the other two and it has a large neutral zone including words like *yellow*,²¹ *yoke*, *yarn*, *yolk*, and *eucharist*. Neutral zone words that (for me at least) risk falling into the penumbra are *Yonkers*, *yak*, and *yam*, which seem a bit silly for purposes of denoting a city, an animal, and a vegetable; see also *Yaphank*, above.

As a phonestheme in Seuss [j-] includes the following core items:

(17) Core [j-]-initial coinages in Seuss

- a. *YOPP*, the cry of help made by a small Who that saves the Whos from destruction (*Horton Hears a Who*)²²
- b. *Yekko*, a beast who “howls in an underground grotto in Gekko” (*On Beyond Zebra*)
- c. *Ying*, a creature with whom it is fun to sing (*One Fish Two Fish*)

But as before, the penumbral examples outnumber them: these include *Yop* (this time a name of a creature, in *One Fish Two Fish*; *Yink*, another creature in *One Fish Two Fish*; and *Yupster*, a place name in *On Beyond Zebra*.

To sum up this section: the patterning of phonesthemes in Seuss’s coinages matches their behavior in real language: we find full-blown core coinages like *Sneedle*, bearing the appropriate meaning; as well as penumbral coinages like *Snumm*, in which the phonestheme provides only expressiveness and style. The third case, namely appearance of the phonesthemic segments without any phonesthetic effect at all, appears to be impossible, since in real life these cases exist only among words that are frequent or learned, neither of which could plausibly be used in a Seuss coinage.

4. Are these speculations on the right track? A statistical test

To return to the main thread, we have sought to explain in general terms Seuss’s coinage practice, and came up with four hypotheses:

- Words that match Seuss’s meter are likely to be Seuss coinages.
- Words that are phonotactically aberrant are likely to be Seuss coinages.
- Words that sound German are likely to be Seuss coinages.
- Words that contain phonesthemes are likely to be Seuss coinages.

²¹ *Yellow* is perhaps penumbral when used to mean ‘cowardly’.

²² As MacDonald (1988:86) observes, the climactic *YOPP* is prepared by the appearance of four evidently-phonesthemic [j-] words (*yapping*, *yipping*, *yo-yo*, *yip*) in the immediately preceding pages. One is reminded of Bergen’s (2004) experimental finding that phonesthemes can be primed.

The statistical model described in §2 was meant to provide the raw material for evaluating these hypotheses in detail. However, that model only tests phoneme sequences as such, and we have not yet tested whether it is really true that it is the phonesthemic status of these sequences, as I have claimed, that is essential. Perhaps Seuss's practice is systematic, but has nothing to do with phonesthemes. Hypothesis-testing in this domain is not straightforward given the notorious subjectivity of phonesthemic analysis.

Hoping to find objectivity, I constructed a second logistic regression model on a different basis. Whereas the previous model was an attempt to scrutinize a great number of potential features, hoping to find the best ones, my second model implements only the proposed phonesthemes found in one single reference source, Marchand (1969); I will call it the Marchand Model.²³ The model is somewhat less complete and accurate than the Full Model given in (3)-(4), but it is arguably objective. Marchand had no axe to grind concerning Seuss, but simply compiled a long list, offering his considered and informed judgment (based on examination of numerous examples), of whether a particular sequence was phonesthemic.

In compiling his list it is clear that Marchand examined all English vowels, all possible onsets, and a great many syllable rhymes.²⁴ In these domains, if Marchand makes no mention of a sequence, it is reasonable to infer that he saw no reason to call it a phonestheme. Unsurprisingly, there is much overlap in the features of the Marchand Model with my Full Model, and to show this, I included the information (page number) of Marchand's discussion of these various sequences in the tables for the Full Model in (3)-(4) above. As before, the complete Marchand Model may be inspected in the Supplemental Materials.

I made two versions of the Marchand Model, coarse-grained and fine-grained. The coarse-grained version implements the intended statistical test. It has just five features, shown in (18).

(18) *Features of the coarse-grained Marchand Model*

PREFER-REAL	This is the intercept term, expressing as before the basic preference for a word being real.
GERMAN	This feature is invoked by all words that invoke one of the specific German features in the Full Model; these are the ones labeled "German" in (3)-(4).

²³ I picked the Marchand source because he tends to be somewhat conservative, refraining from seeing phonesthemes everywhere he looks. Magnus (2001) and Bolinger (1965) are, conceivably, correct in claiming that phonesthemes are omnipresent, but if this view is true then the hypothesis "Seuss used phonesthemes in coining words" becomes trivial and not worth checking.

²⁴ Indeed, deviating from his normal practice, he seeks a phonesthemic interpretation for *every* vowel, so there was no point in including them in my Marchand Model. Marchand also covers a few singleton final codas, but they did not improve the accuracy of the model and I omit them here.

PHONOTACTIC	This feature is invoked by all words that are phonologically ill-formed or near-ill-formed, invoking one of the features labeled as such in (3)-(4). ²⁵
METRICAL	The sum of the violations of the two meter-related constraints discussed in Appendix C.
MARCHAND-ONSET	This feature is invoked by all words that begin with a Marchand-mentioned onset phonestheme.
MARCHAND-RHYME	This feature is invoked by all words that end with a Marchand-mentioned rhyme phonestheme.

I fitted the coarse-grained Marchand Model to the same data as before, again using the `BayesGLM()` package in R. The constraint weights and significance values that were calculated are given in (19).

(19) *Result of fitting the Coarse-Grained Marchand Model*

<i>Feature</i>	<i>Weight</i>	<i>P-value</i>
PREFER-REAL	-4.87	< .0001
GERMAN	3.45	< .0001
PHONOTACTIC	4.91	< .0001
METRICAL	1.23	< .0001
MARCHAND-ONSET	0.92	< .0001
MARCHAND-RHYME	1.15	< .0001

The model, being so coarse, is far less effective than the model of (3)-(4) in predicting Seussian status; see Appendix A for details. The key point of the model is that it permits Marchand's independent testimony to bear on the question of whether Seuss's practice is indeed phonesthemic. The results of the model are that Germanness, phonotactic ill-formedness, and metrical appropriateness all test significant as factors for predicting Seussian status. In addition, phonesthemic status, for sequences identified as such by Marchand, also matters; although the constraint weights may be lower than those for Germanness and phonotactics, the number of words covered is considerably larger.

To obtain a more detailed look, I also ran a fine-grained version of the Marchand Model that separates out all the Marchand-mentioned features (there are 44 for onsets and 116 for rhymes). The result, available in the Supplemental Materials, demonstrates that most of the work of

²⁵ Of course the German sequences are themselves mostly ill-formed in English, but for clarity I excluded them from the scope of my Phonotactic feature, which covers only the remaining ill-formed cases.

predicting Seussian status is being done by a fairly small subset of Marchand's features; only 21 of 160 meet the criterion of bearing a weight of at least 1 and receiving a p -value $< .001$.²⁶

The upshot of these studies, I believe, is as follows. If we agree to take Marchand as an impartial witness for phonesthemic status, then it seems almost certain that Seuss is using phonesthemes when he coins words. Further, Seuss is making use of only a modest subset of Marchand's phonesthemes. There are at least two possible reasons for this. First, Marchand may have been overenthusiastic in positing phonesthemes (I tend to think so, particularly among the onsets). Second, Seuss may have been making a perhaps-unconscious artistic decision, choosing his favorites from a larger available inventory.

5. Conclusions

A verbal artist, particularly a popular artist, must rely on phonological resources he shares with his reading community. This dictum is confirmed by Seuss's coinage practice. First, native speakers of English internalize a detailed phonotactics of their language, which leads them to be amused by novel words like *Thneed*. Speakers also have some ability to internalize phonotactic principles of languages they don't speak but find accessible, and hence can be entertained by novel pseudo-German words like *Schlottz*. Lastly, they have internalized a system of phonesthemes, which gives them the ability to appreciate novel phonesthemic words. Just like real-life phonesthemic words, coined ones may either include the semantic component of the phonestheme, as in *Sneedle*, or exclude it, with the phonestheme offering only a sense of style, as in *Snumm*.

It goes without saying that the rigor of the research reported in this article would be increased by extensive experimentation, in the research tradition of, e.g., Fordyce (1988), Hutchins (1998), and Bergen (2004). We would like to know more about which proposed phonemes are actually internalized by native speakers, what meanings they are assigned, whether my proposed "penumbra" (§3.4.1) is psychologically real, and (on a different topic) the extent to which older American English speakers (the original Seuss audience) have internalized the phonotactics of German (§3.3.1). Since my account depends on the ability of people to learn the stylistic affiliation of particular linguistic entities, we are also in need of a motivated theory of how this is done.

Lastly, it might also be useful to carry out studies comparing Seuss's use of phonesthemes with that of other word-coiners — in literature, in ordinary life (novel words-without-etymology), and in industry (see Wong 2014, and the Pokémon research cited in §1). I imagine that such study would find considerable variation. While Seuss's choices were principled, I suspect that they access only a subset of the possibilities offered by the resources that the language offers — this is what the fine-grained Marchand Model (§4) suggests. My conjecture reflects the view (§3.4) that word-coinage is a folk art: within the limits of what their language makes available, verbal artists can make choices.

²⁶ These 21 are listed as follows in descending order of weight: [z-], [sn-], [-ap], [-ʌmp], [-ʌd], [-ʌm], [-ʌb], [bl-], [-u:n], [gl-], [-ek], [skr-], [-u:], [g-], [j-], [kw-], [gr-], [w-], [fl-], [b-], [dʒ-].

Appendix A: Performance of the models compared

The metrics given here are explicated in, e.g. Johnson (2008).

	<i>Full</i>	<i>Marchand-Coarse</i>	<i>Marchand-Fine</i>
<i>Number of features</i>	46	6	167
<i>Log likelihood of the data corpus</i>	-1437.1	-1882.7	-1581.8
<i>Akaike Information Criterion</i> (lower is better)	2966.1	3777.4	3497.7
<i>Average probability of a Seuss word</i>	.211	.079	.151
<i>Average probability of a real word</i>	.019	.023	.021

Appendix B: Coinages homophonous with real words

I excluded from the regression analyses the Seuss 66 “coinages” that exist as real words, but are used in context as novel. For instance, *FlummoX* is used in Seuss not as a verb but as a noun, the name of an imaginary creature in *If I Ran the Circus*. From the viewpoint of the Full Model developed in §2, these “semi-coinages” appear to have an intermediate status, as the figures in (20) 202020indicate.

(20) Average Seussian probability of novel Seuss words:	0.211
Average Seussian probability of real words employed as imaginary words:	0.080
Average Seussian probability of real words employed as such:	0.019

The ten real words with highest Seussian model probability are *Zipp* (0.722), *FlummoX* (0.489), *Krupp* (0.338), *guff* (0.329), *Snide* (0.264), *Fuddle* (0.227), *Quibble* (0.222), *duff* (0.179), *Snell* (0.167), *Gusset* (0.15).

Appendix C: The coinage principles motivated by meter

I discuss in this appendix two principles of Seuss coinage that are based on the fact that he would naturally favor coinages that fit well with his favorite meter. This is anapestic tetrameter, which in its base form is /x x X x x X x x X x x X/. Seuss took care to write this meter with strict adherence to syllable count; i.e. he sought never to invoke the license (common among his parodists) of substituting a binary for a ternary foot. Hence, words like *Bippo-no-Bungus* or *Motta-fa-Potta-fa-Pell*,²⁷ which have the sequence 'σ ǝ ǝ 'σ, serve a useful metrical purpose. They are fairly common in the Seuss oeuvre, despite his general dispreference for long coinages ((3a)).

A special case is found in words that include the medial sequence [əmə], as in *Katta-ma-side*, *Yuzz-a-ma-Tuzz*, and six similar instances. This sequence appears in real English words, usually vernacular; the cases known to me are *razzamatazz*, *rigamarole*, *tacamahac*, *Fishamajig*,

²⁷ Seuss often includes hyphens in the spelling of his long coinages. I suspect these are not intended to denote pseudo-morphological structure, but are only intended to make the long words easier to read. Teuber (2018) suggests that hyphen placement also aids the reader in assigning stress; e.g. *Va-Vode* comes across more clearly as finally-stressed than *Vavode* would be.

whatcha-ma-callit, *thingamajig(ger)*, *thingamabob*, and *Kalamazoo* (the last three appear in Seuss). The morphological status of [-əmə-] is obscure to me, but it does seem to be productive (e.g., *rigamarole* evolved from earlier *rigmarole*; *OED*); perhaps it is a phonestheme. In any event, [-əmə-] provided Seuss with a source for words that are both colloquial and metrically facilitating.

References

- Baayen, R. H., R. Piepenbrock, and L. Gulikers (1995) *CELEX2 LDC96L14*. Web Download. Philadelphia: Linguistic Data Consortium.
- Bergen, Benjamin K. (2004) The psychological reality of phonaesthemes. *Language* 80:290-311.
- Bolinger, Dwight (1965) *Forms of English: Accent, morpheme, order*. Cambridge, MA: Harvard University Press.
- Chambers, W. Walker and John R. Wilkie (1970) *A short history of the German language*. London: Methuen.
- Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. (2011) Explaining sonority projection effects. *Phonology* 28:197-234.
- Davis, Stuart (1991) Coronals and the phonotactics of nonadjacent consonants in English. In Carole Paradis and Jean-François Prunet, eds. *The special status of coronals: Internal and external evidence*, pp. 49-60. San Diego, CA: Academic Press.
- Firth, J. R. (1930) *Speech*. London: Ernest Benn.
- Eisiminger, Sterling (1981) Etymology unknown: Toward a master list of words of obscure origin. *American Speech* 56:146-148.
- Fordyce, James (1988) *Studies in sound symbolism, with special reference to English*. Ph.D. dissertation, UCLA, Los Angeles, CA.
- Hammond, Michael (1999). *The phonology of English: A prosodic optimality-theoretic approach*. Oxford: Oxford University Press.
- Hayes, Bruce (2016) Comparative phonotactics. *Proceedings of the 50th Meeting of the Chicago Linguistic Society*, 265-285.
- Hayes, Bruce and Colin Wilson (2008) A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379-440.
- Hutchins, Sharon Suzanne (1998) *The psychological reality, variability, and compositionality of English phonesthemes*. Ph.D. dissertation, Department of Psychology, Emory University.
- Jespersen, Otto (1922) *Language: Its nature, development, and origin*. London: Allen & Unwin.
- Johnson, Keith (2008) *Quantitative methods in linguistics*. Malden, MA: Blackwell.
- Joseph, Brian (1994) Modern Greek [ts]: Beyond sound symbolism. In Hinton, Leanne, Johanna Nichols, and John J. Ohala, eds., *Sound Symbolism*. Cambridge: Cambridge University Press, pp. 222-236.
- Kawahara, Shigeto (2020) Sound symbolism and theoretical phonology. *Language and Linguistics Compass* 14:e12372.
- Kawahara, Shigeto (2021a) How Pokémonastics has evolved: Ver 1.0. <http://user.keio.ac.jp/~kawahara/pdf/PokemonasticsOverviewVer1.0.pdf>
- Kawahara, Shigeto (2021b) *Phonetic bases of sound symbolism: A review*. <https://psyarxiv.com/fzvsvu>.
- Keyes, Ralph (2021) *The hidden history of coined words*. Oxford: Oxford University Press.

- Lathem, Edward Connery (2000) *Who's who & what's what in the Books of Dr. Seuss*. Hanover: Dartmouth College.
- MacDonald, Ruth K. (1988) *Dr. Seuss (Theodore Seuss Geisel)*. Woodbridge, CT: G. K. Hall & Company.
- Magnus, Margaret (2001) What's in a word? Studies in phonosemantics. Ph.D. dissertation, Norwegian University of Science and Technology, Trondheim.
- Malkiel, Yakov (1990) *Diachronic problems in phonosymbolism*. Amsterdam: J. Benjamins.
- Marchand, Hans (1969) *The categories and types of present-day English word-formation*. 2nd edition. Munich: Verlag C. H. Beck.
- Morgan, Judith and Neil Morgan (1995) *Dr. Seuss & Mr. Geisel: A biography*. New York: Random House.
- Liu, Nelson F., Gina-Anne Levow, and Noah A. Smith (2018) Discovering phonesthemes with sparse regularization. In *Proceedings of the Second Workshop on Subword/Character Level Models*, New Orleans, Louisiana, June 6, 2018. Association for Computational Linguistics, pages 49–54.
- Nilsen, Don L. F. (1977) Dr. Seuss as grammar consultant. *Language Arts* 54:567-574.
- Oh, Y., J. Needle, Simon Todd, Clay Beckner, Jennifer Hay, and Jeannette King (2020) Non-Māori-speaking New Zealanders have a Māori proto-lexicon. *Scientific Reports* 10:22318.
- Otis, Katya and Eyal Sagi (2008) Phonaesthemes: A corpus-based analysis. *Proceedings of the Annual Meeting of the Cognitive Science Society* 30:65-70.
- Pentangelo, Joseph (2020) Phonesthetics and the etymologies of *blood* and *bone*. *English Language and Linguistics* 25:225–255.
- Prince, Alan and Paul Smolensky (1993/2004) *Optimality Theory: Constraint interaction in generative grammar*. Technical report, Rutgers University Center for Cognitive Science. [Published 2004; Oxford: Blackwell]
- Schmidtke, David S., Markus Conrad, and Arthur M. Jacobs (2014) Phonological iconicity. *Frontiers in Psychology* 5:80.
- Shih, Stephanie S., Jordan Ackerman, Noah Hermalin, and Sharon Inkelas (2018) Pokémonikers: A study of sound symbolism and Pokémon names. *Proceedings of the Linguistic Society of America* 3.42:1-6.
- Teuber, Daniel (2018) An analysis of the nonce-words of Dr. Seuss. *Journal of Osaka Sangyo University* 34:43-69.
- Wescott, Roger (1980) “Zazzification” in American English slang, in his *Sound and sense: Linguistic essays on phonosemic subjects*. Lake Bluff, IL: Jupiter Press, pp. 391-393.
- Wong, Andrew (2014) Branding and linguistic anthropology: Brand names, indexical fields, and sound symbolism. *Practicing Anthropology* 36:38–41.