

Joshua T. Katz

Exercises in wile

Abstract: At Princeton University, I have for the past decade occasionally taught a course – a so-called freshman seminar – on the history and practice of wordplay: “Wordplay: A wry plod from Babel to Scrabble”. Zany, rigorous, and popular, the class gives students with interests from literature to mathematics the opportunity to explore the ludic side of language through a combination of three sorts of activities: the reading of primary literature, the consideration of secondary scholarship, and the regular and active creation of new instances of wordplay. The subject is not frivolous. With all due respect to colleagues who study “core” phenomena, I contend that pushing against what one might think of as the margins of language is fascinating in itself as well as an under-utilized but effective way to introduce people to the subject of linguistics. This paper attempts to give the flavor of this offbeat seminar, among other things by describing series of exercises designed to highlight some of the more simultaneously striking and easily exploitable orthographic, phonological, and lexical peculiarities of English. My hope is that this will be seen as a practical contribution to what the call for papers for the conference in Trier referred to as “systematic and analytical approaches to wordplay, its forms and functions”.

Keywords: alphabet, Christian Bök, Doug Nufer, emoji, frequency effects, Georges Perec, linguistic exercises, lipograms, marginal phenomena, Oulipo, pedagogy, Scrabble, sounds, wordplay, Zipf’s law

1 Introduction: From Babel to Scrabble

About a decade ago, one of my extracurricular pursuits since childhood, playing around with words, became an adult academic interest as well. While writing a couple of papers on elaborate verbal games in the works of the Roman poet Vergil (70–19 BC) – one that involves reading a word backwards at the start of the *Georgics* (Katz 2008) and another, more speculative, on a would-be acrostic in Book 4 of the *Aeneid* (Katz 2007) – I realized that it would be helpful to have a broader and deeper view of wordplay: what the general designation “wordplay” encompasses in Latin, in English, and in language in general; what forms of wordplay have existed, flourished, and been celebrated across time and space; and how puns, palindromes, word squares, and the like can be and have been devised, interpreted, and used for very different reasons and with very different

effects, a matter that called for perspectives beyond my accustomed historical, literary, and linguistic ambit, namely anthropological, psychological, sociological, religious, mathematical, and visually artistic. The problem was that there did – and still does – not appear to be a good general introduction to wordplay¹, though bookstores sell innumerable puzzle books of all sorts and library shelves and online repositories hold plenty of specialized studies on topics from autograms to tautograms and from Ancient Near Eastern oneirocritical wordplay to teenagers’ ludic uses of the latest electronic devices.² As every teacher knows, one of the best ways to learn about a subject is to give a course, and so in the fall of 2009, I decided to offer a so-called freshman seminar, one of a significant number of typically quirky classes, capped at fifteen students apiece, that Princeton University encourages first-year undergraduates to apply to take so that they begin their collegiate career by learning about something interesting and offbeat in an intimate setting under the tutelage of a professor they get to know well. The final sentence of the final footnote of a programmatic paper published just as the term was beginning – it is titled simply “Wordplay” (Katz 2009)³ – reads as follows: “My goal is to come out of the class knowing much more than when I went in” (105, n. 90).

1 In my experience (in the class described below), linguist David Crystal’s book *Language Play* (Crystal 1998b) does not fit the bill; Augarde (2011) presents many nice facts about subjects from “spurious words” to contronyms but could not be used as a textbook. Cook (2000), which concentrates on (but goes well beyond) the role of play in language pedagogy, remains a standout. Borgmann (1965) and Eckler (1996) provide unparalleled entertainment.

2 Hofstadter (1985: 64–65, 68–69, and 391–392) gives the classic, albeit brief account of what the English engineer and recreational mathematician Lee Sallows has called an autogram (see http://www.leesallows.com/index.php?page_menu=Self-referential%20stuff&pagename=In%20Quest%20of%20a%20Pangram, accessed 29 July 2018, with Sallows 1985), a self-referential text that enumerates its own letters and/or other characters; I do not know of any special scholarly literature on tautograms, but they are the visual analogue of alliteration, on which there is no shortage of research (see, briefly, Adams and Cable 2012). For ludic oneirocriticism, see Noegel (2007). Unsurprisingly, traditional published books cannot keep up with technological changes, to the point that my eighteen-year-old students in 2009 deemed *Txtng: The Gr8 Db8* (Crystal 2008) embarrassingly out-of-date; the value of this work to their counterparts now, who use iPhones rather than flip phones and emoji rather than emoticons, lies largely in its documentation of a lost world.

3 Katz (2009: 79–84) provides a brief account of another newly spotted instance of Latin wordplay, this time in Cicero, to add to the ones described in the papers cited above in the text. For further accounts of wordplay in the ancient world, see Katz (2013b) on, *inter alia*, a new acrostic in Vergil and Katz (2016a) on a new acronym (see also, briefly, Katz 2014a and 2014d); Katz (2013c and 2015) on Ferdinand de Saussure’s “anagram notebooks”; and Katz (2013a) on

I achieved this goal, and I hope, too, that I can accurately claim that the students learned a lot in “Wordplay: A wry plod from Babel to Scrabble”, as I called our joint sallies into the history and practice of sometimes quite remarkable verbal gymnastics. The title, besides containing an ironic anagram (*wordplay ~ a wry plod*)⁴, is meant to point simultaneously, in part through an oblique nod to the homophony of the words *Babel* and *babble*, to the role of less-than-straightforward linguistic undertakings both across the sweep of human history (from the story in *Genesis* 11:1–9 to a modern board game) and in the history of each individual human (from the babbling of infancy to a world in which winners and losers are decided through the casting of lots marked with numerically assigned letters).⁵ By this point, I have taught the class four times (2009, 2011, 2012, and 2015), and on each occasion it has been a happy success, drawing a substantially larger number of applications from incoming freshmen than there were places available. Furthermore, the “plodders” who end up enrolling have tended to be just the extraordinarily bright and extraordinarily zany sorts I want: humanists who are not scared of formal systems and mathematicians who are interested in language and literature. The course has attracted local and national attention: in 2010, the *Princeton Alumni Weekly* ran a story (“Word freaks”) about the initial offering; in 2011, it was ranked #2 on *Mental Floss*’s list of “22 fascinating and bizarre classes offered this semester” (the top spot went to a class on Lady Gaga at the University of South Carolina) and #7 on *The Daily Beast*’s list of the eighteen “Hottest college courses”; and *Time* magazine in 2015 profiled it among the “11 bizarre college courses we actually want to take”.⁶

various *jeux*, in Greece and elsewhere, that involve both sounds and script. Katz (2014b and 2014c) gives quick accounts of Greek puns and riddles.

4 I am indebted to Liesl Yamaguchi for the delightful French translation “Jeu de mots : mode juste de Babel au Scrabble”. For German, I propose “Wortspiel: List Power von Babel zu Scrabble”.

5 In my dialect of English, *Babel* and *babble* are homophonous: [ˈbæbəl]. However, many speakers pronounce the former rather as [ˈbeɪbəl]. In any case, while the ultimate derivation of *Babel/Babylon* is unknown, folk etymology – for whose status as an intermediary between scientific etymology and wordplay I have argued in Katz (2010a, 2010b, and 2016b) – gives it the Akkadian interpretation *bāb ili* ‘gate of God’ and (of particular relevance here) accounts for the biblical story, in which what happens at Babel is explained through the Hebrew verb *balal* ‘mix, confuse, confound’. In a brief paper titled “From Scrabble to babble: Reflections on language attitudes and language play”, Crystal (1998a) argues for the social utility of studying wordplay.

6 See respectively Noden (2010) and the follow-up piece “Wordplay winners” (2010); Conradt (2011); “Hottest college courses” (2011); and Waxman (2015).

Those who have taken the course, the most recent catalogue description of which may be found at <https://www.princeton.edu/pub/frs/ay201516/fall-courses/index.xml#compfrs123> (accessed 29 July 2018), have engaged in rigorous explorations of ludic language through a combination of three sorts of activities: (1) the reading of primary literature (e.g., poems, stories, and novels by Raymond Queneau, Vladimir Nabokov, Georges Perec, Walter Abish, Paul Muldoon, Christian Bök, and Mark Dunn); (2) the consideration of secondary scholarship on Oulipo, spelling bees, the art of the crossword, graphic wordplay in languages with non-alphabetic writing, and much more; and (3) the regular and active creation by all participants, followed by group discussion, of new and often delightfully inventive instances of wordplay.⁷ Of course, in order to be able to discuss properly the aesthetic pleasures and societal roles of wordplay, one needs – or should ideally wish – to have a definition of the term, to be able to enumerate its formal features.

2 What is wordplay?

Defining our term turns out to be difficult, and not merely because it is hard to give a precise account of either “word” or “play”.⁸ For one thing, if there is a

⁷ There is a “Princeton dimension” to the course. Paul Muldoon, whose 1990 poem “Capercaillies” (see, conveniently, Muldoon 2001: 198–199) I assign for the first class (along with Nabokov’s classic 1951/1959 short story “The Vane sisters”) has come to talk to the students, and we also generally consider works by at least two other colleagues of mine: David Bellos’s English translations of and scholarship on Perec (e.g., Bellos 1995) and biophysicist Bill Bialek’s foray into alphabetic entropy, which he illustrates in part through the totality of four-letter words present in the corpus of Jane Austen (Stephens and Bialek 2010; see also Bialek’s chagrined comment re paper #127 at http://www.princeton.edu/~wbialek/publications_wbialek.html, accessed 29 July 2018). Trophy-winning spellers in the student body and a Scrabble champion on the staff also come by to lend their perspectives and challenge us with a smile to best them, which few can do.

⁸ For attempts to come to terms with the essence of wordplay, see the volumes already published in the series in which the current collection appears, “The Dynamics of Wordplay”: Zirker and Winter-Froemel (eds.) (2015), Winter-Froemel and Zirker (eds.) (2015), and Knospe, Onysko, and Goth (eds.) (2016); Winter-Froemel (2016, esp. 37–42) and Thaler (2016) are especially valuable. The definition of “word” need not overly trouble us (though see sections 5 and 6, especially fn. 53): “language play” would be a better designation of the phenomenon under discussion than “wordplay”. As for play, Johan Huizinga in his classic account “[s]um[s] up the formal characteristics of play” as follows: “[W]e might call it a free activity standing quite consciously outside ‘ordinary’ life as being ‘not serious’, but at the same time absorbing

sentence, or some other utterance, in English – or, as far as I know, in any other language – that is *not* an example of wordplay, I have yet to imagine it. Or perhaps I should say that *could not* be an example, for wordplay is in part a matter of context. Take, for instance, the sentence *The cat is on the table*. As such, nothing suggests that this is anything other than an ordinary utterance, with no ludic dimension. But it would be easy to take it and add a rhyme, thereby injecting a clear element of play – and easy, too, to continue to a full-blown poem: *The cat is on the table; / Her fur resembles sable. / Majestic, she rules – / With gules*. (The fact that this is *bad* poetry is beside the point.) Alternatively, one could play with *The cat is on the table* by highlighting that all the words but the last are monosyllabic (*The cat is on the table; a ball lies in the gutter; no girl sings at John's party*); or that the second word begins with the letter *c* (*The cat is on the table; so are the cod, the cleat, the cereal, and the chasuble*) or the sound [k] (*The cat is on the table; so are the kit, the khat, the qanun, and the chemical*); or that the first and fifth words are identical (*The cat is on the table; fish that like other fish swim; I am happy because I eat*); etc. etc. etc. What makes all of these cases of wordplay – and what makes puns, autograms, and games of Scrabble cases as well – is that each of them not only involves some degree of self-reflexivity, pointing in its very form to its status as language, but elevates form to the point of content and sometimes even beyond.⁹ The element of play here is clear enough – play is deliberately self-conscious, “standing”, as Huizinga (1949: 13) puts it, “quite consciously outside ‘ordinary life’” (see fn. 8 above) – but a bit more time should be spent on the meaning of “language”.

Suppose we imagine language as a ball. Most linguists are interested principally in “core” phenomena, that is to say, in what lies at and near the center of the ball; a few, however, find it interesting to consider more “marginal” material, what in German are called *Randphänomene*, figuring that any good theory has to account for everything that counts as language, not just what is usual.¹⁰ This leads to an obvious question: what does *not* count as lan-

the player intensely and utterly. It is an activity connected with no material interest, and no profit can be gained by it. It proceeds within its own proper boundaries of time and space according to fixed rules and in an orderly manner. It promotes the formation of social groupings which tend to surround themselves with secrecy and to stress their difference from the common world by disguise or other means” (Huizinga 1949: 13); see also, e.g., Cook (2000: 97–120 and *passim*).

⁹ I repeat this definition of wordplay – or, perhaps better put, description of the essential character of the phenomenon – nearly verbatim from Katz (2009: 100–101 and 2013b: 3).

¹⁰ Gordin and Katz (forthcoming) discuss what we term “non-intrinsic philological isolates”, i.e., unique – and hence marginal – forms of language, some of them involving significant

guage? Whereas in the last paragraph I endeavored to show that it is essentially impossible to give a good answer to the question, “What does *not* count as wordplay?”, there is, I believe, a useful way in the present context to say what is not language: it is what lies outside the ball. To take an infamous example from Chomsky (1957: 15): while *Colorless green ideas sleep furiously* (sometimes called the “Chomsky sentence”) is certainly a curious example of English, nowhere near the core from a semantic and pragmatic point of view, it is in fact – and undeniably – English; by contrast, its inverse, *Furiously sleep ideas green colorless*, though it contains the very same five words and thus appears to be language, is “word salad”. In the second “sentence”, the five words are given in an order that not only makes them (arguably) not an example of English but from some perspectives, including (arguably) the playful one, places their collocation outside the ball of language entirely.¹¹

The relevance of this to the present paper is that the instances of wordplay I go on to highlight are extreme in the sense that they toy – and not merely in passing – with the ball’s boundary without, crucially, going beyond (compare Katz 2013b: 3: wordplay as “*extreme language*”). Whenever I teach the seminar on wordplay, I start by getting my students to see how difficult the term is to define and then move straight to what I sometimes call “language on the verge”, that is to say, to *Randphänomene*, coaching them how to skate along the edge by putting them through series of exercises that revolve around the evaluation and production of original examples of verbal art that are at the same time especially peculiar and – because they adhere to a single unusual but easy-to-articulate principle, or to a small set of them – in one way really quite simple to understand. To quote the Oulipian Georges Perec (1936–1982), arguably the greatest literary producer of work based on unusual principles that are at least

wordplay, that could be used outside the confines of the literary works in which they made their debut but generally are not.

¹¹ Recognizing that I am probably pushing the idea too hard, I have included two parenthetical “arguably”-s. To be sure, there are perspectives, especially neurological and psychiatric ones, in which word salad – a subject Ray Gibbs treated with verve in his talk at the conference in Trier – may still be inside the ball: a patient who produces the string *Furiously sleep ideas green colorless* will be treated as having a language disorder, not as not having language. As for the considered use of *Furiously sleep ideas green colorless*, I remain doubtful that this has much of anything to tell us about the ludic function of language, but see Erard (2010, esp. 424–425), who prints one of two poems by the Anglo-Scottish linguist and philologist Angus McIntosh that include it. The literature on Chomsky’s two sentences from everything from a literary to a statistical point of view is large: in addition to Erard’s essay, see also, e.g., Jahn 2002 (as well as Katz 2013b: 4, n. 10 and the works cited there).

sometimes easy to articulate, “Au fond je me donne des règles pour être totalement libre” (Perec 1977).¹² Huizinga (1949: 13) has called play “a free activity [...] [that] proceeds [...] according to fixed rules” (see fn. 8 above), and one of the leading lessons of my seminar on specifically verbal play is that strict adherence to rules or principles – or, as they are often called, constraints – is, paradoxically, often liberating, with the surprising sense of freedom yielding, in the best cases, extraordinary effects.

The “exercises in wile”¹³ that follow are of three kinds: alphabetic, phonological, and lexical.¹⁴ Though based on English, many are adaptable to other languages (some more easily than others);¹⁵ I have selected these specific ones for inclusion because they work neatly together, as I hope will become clear. It should go without saying that alphabetic exercises do not apply to languages that lack a written tradition or make standard use of an orthographic practice very different from the ABCs; other forms of writing – abugidas, syllabaries, logographic systems, etc. – engender further, often wonderful forms of word-play, discussion of which, however, goes beyond the confines of this paper.

12 Levin Becker (2012: 13) writes that Perec’s comment “has come to epigrammatize the Oulipo as a whole”.

13 An homage, of course, to *Exercises in Style*, Barbara Wright’s 1958 English translation of the 1947 collection of stories *Exercices de style* by Raymond Queneau, who would go on to co-found Oulipo in 1960. Only after submitting my abstract for the conference in Trier did I discover that Stump (2003) had anticipated me in the title “Exercises in wile”.

14 I start with alphabetic exercises rather than ones concerned with sound because most people in a highly literate society are far more attuned to the nurtured behavior that is spelling than to the subtleties of the natural sonic environment; compare fn. 47 below.

15 I cannot resist pointing out a peculiarity that manifested itself in the multilingual environment of Trier, where the talks at a conference at a German university were in English and French. Because the point values of Scrabble tiles depend on letter frequency (a subject I discuss in section 3) and because languages do not by any means have the same distribution of letters, one cannot reasonably play Scrabble with a set designed for the wrong language. The poster for the conference featured a photograph of a Scrabble rack with the word *words* spelled out as follows: W₄O₁R₁D₂S₁. This is English (French has W₁₀ and German has W₃, O₂, and D₁). The inside of the program, however, had a photograph of W₂O₂R₁D₁P₄L₂A₁Y₁₀ and J₆E₁U₁X₈D₁E₁M₃O₁T₂S₁. Amusingly, both the English and the French here are spelled with German tiles – and not just any German tiles but ones from a set that predates 1989/90, after which W₂ was replaced by W₃!

3 The alphabet

Consider the following lipogrammatic exercise:

- (1) a. Choose one of the five vowels in our alphabet (*a, e, i, o, or u*) and compose an English sentence of at least seven words in which the only vowel used is that one. (You may use any consonants you wish.)¹⁶

At the conference in Trier, two colleagues in the audience responded quickly, and from my perspective most satisfyingly, to the challenge: Dirk Delabastita produced *A bat has had a bad fall* and Eline Zenner *Ernest the ewe gets the best bet yet* – the latter after first having come up with *The earnest ewe gets the best bet yet*, which contains the disallowed word *earnest*. (This sort of mistake is interesting in its unsurprisingness: experience shows that even highly educated people of exquisite linguistic sensibility frequently make basic errors, a sign that the task they are asked to perform falls well away from the center of the ball of language.) After years of recording how students and colleagues respond to the exercise, I can state unhesitatingly that answers with <a>'s or <e>'s abound, sometimes there are <i>'s or <o>'s, but nearly no one has led off with <u>'s;¹⁷ the reasons for this surely have to do with the strong tendency of people to begin with an article – *a(n)* or *the* – and the fact, not unconnected, that <e> and <a> are especially frequent letters but <u> substantially less so.¹⁸ When one considers the most frequently used letters in any large corpus of English – the conventional order begins with the happily pronounceable ETAOIN SHRDLU – one immediately notices the extraordinary fact that four of the five vowels are among the five most common letters and that even the outlier, <u>, ranked

16 Aside from being the “magical number”, there is nothing special about seven; anyone wishing to produce answers to this and other exercises with only five or six words is perfectly welcome to do so. Readers may also wish to attempt the exercise using none of the five vowels and only y's. As for employing y's in addition to one of the standard five vowels, I suggest doing the exercise first without and then with them – and concentrating on what difference this makes.

17 A standard linguistic convention is to use angled brackets to indicate letters of the alphabet; by contrast, square brackets, which we will see in section 4, indicate pronunciation.

18 It is also the case that people tend to think of short words with <u> as vulgar: *cum, cunt, fuck, slut*, etc. Reading Bök (2005: 79) will reinforce this impression (see fn. 29 below), though the poem “Stuck-up gulls must trust dumb ducks” in JonArno Lawson’s wonderful collection for children *A Voweller’s Bestiary from Aardvark to Guineafowl (and H)* (Lawson 2008: 16) belies it (it does, however, contain the naughty phrase *thumps gull’s rump*).

twelfth, is in the top half of the alphabet.¹⁹ No wonder, then, that while it is not that hard to produce sentences that use only <e>, <a>, <o>, or <i>, precisely because they are all so ordinary, the devilish catch is that the exercise at the same time eliminates most of the other letters that one would normally employ.

To return to the answers I am used to receiving: most are made up largely of vocabulary items that are monosyllabic, semantically basic, or both (in the Trier sentences, all but *Ernest*, which as a name falls into rather a special category); noteworthy (though, as I have just indicated, not at all peculiar) is that Dirk and Eline, besides attempting in the first place to begin with *a* and *the*, respectively, each made double use of an article. Furthermore, responses very often have an animal as their subject (*bat* and *ewe*, but also *cat*, *rat*, *elk*, *pig*, *ox*, etc.)²⁰, which suggests that the exercise triggers a desire on the part of adults to return to the state of *1 Corinthians* 13 (“When I was a child, I spake as a child,...”) and, indeed, to revel in ludic activities associated with primary school.²¹

There is also the matter of meaning. I applauded Eline’s switch from the adjective *earnest* to the proper name *Ernest*, but because the exercise takes our attention away from what most of us use language for most of the time, namely communicating sense, it was a minute or so before some of us, including me, realized that there was something absurd about *Ernest the ewe* or, as Eline then wryly put it, *Ernest the genderbender ewe*.... There are pragmatic issues with Dirk’s sentence as well, though they are more subtle: both the indefiniteness of

19 Letter frequency used to be a matter of importance to more than wordplayers and specialists in cryptography: when manual typesetting was still the way in which most books were made, type was metal and cumbersome, and it would have been absurd for anyone who was producing texts in English to have to hand anything like as many <z>’s as <a>’s. David Loeb Weiss’s short documentary film “Farewell, etaoin shrldu” (available at, e.g., <https://www.nytimes.com/video/insider/100000004687429/farewell-etaoin-shrldu.html>, accessed 29 July 2018) records the night of 1 July 1978, when the *New York Times* was set for the last time in hot type, largely on Linotype machines. There is a considerable body of secondary literature on alphabetic statistics in English and elsewhere; Norvig (2012), using data from Google Books to build on Mayzner and Tresselt (1965), one of a number of classic papers by Mayzner et al. that report data gathered by means of IBM punch cards, is an excellent place to start – for letter frequency in general, the relationship between letter frequency and position in a word, and much else besides, including word frequency, a topic I consider at some length in section 5. Unfortunately (?), Norvig reports that the correct order is actually ETAOIN SRHLDCU, with <u> ranked thirteen.

20 See Lawson (2008), with fn. 18 above.

21 The word *ludic* comes from *ludus*, the Latin word for both ‘game, sport, play’ and ‘place of instruction or training, school’; for the connection between the two senses, see Katz (2013b: 24–25).

the subject and the present perfect tense make *A bat has had a bad fall* odd, whether decontextualized or as the opener of a (children's) tale.²² We might expect *The bat had a bad fall*²³ – but this is not licit in an environment in which a sentence or story with *a*'s cannot have definite articles, just as one with *e*'s cannot have indefinite articles and one with *i*'s, *o*'s, or *u*'s cannot have any articles at all.

Lipograms are thus not so easy to produce, and all who try their hand at them will do well to ponder the following question:

- (1) b. What problems do you encounter?

Above all, exercise (1a) forces highly competent speakers, writers, and readers to come to terms with the different and sometimes dramatic domino effects that result from cleaving to one or another seemingly innocent constraint. For example, no story can talk about *you*, though an *i*-story can talk about *I* (and *him* and *it*), an *e*-story about *me* (and *he*, *she*, *her*, *we*, *they*, and *them* – and, if one wishes, *thee*), and a *u*-story about *us*; past tenses are easiest in an *e*-story, thanks to the ending *-ed*, while present participles (in *-ing*) are possible only in an *i*-story; a story that asks *who* cannot also ask *what* or *where*; no story can give an explanation with *because* or *since*; the conjunctions *and*, *or*, and *but* are impossible except in *a*-, *o*-, and *u*-stories, respectively, only the middle of which can have *no* and *not*; and the preposition *by* can appear anywhere, *near* and *out* can appear nowhere, and *at*, *between* (and *except*), *in* (and *with*), *for* (and *down*, *of*, *on*, *onto*, and *to*), and *up* cannot be intermingled.

I have been talking about stories rather than simple sentences because even more serious problems with exercise (1a) arise when one takes things a step further:

- (1) c. Now try to write a follow-up sentence.

A follow-up sentence has to follow up, that is to say, has to continue the theme established by the initial one. But because people rarely respond to the first exercise by producing a sentence in which semantics is paramount, they are

²² It would work with an elaborate set-up: a passerby stumbles on an unknown chiropter with broken bones and calls emergency services, saying, "A bat has had a bad fall. Send an ambulance immediately!" (This is to say nothing of the interpretation in which *fall* is synonymous rather with *autumn*.)

²³ The sentence *A bat had a bad fall* has only six words, which is why Dirk felt he had to add the extra *has*.

usually thoroughly flummoxed by this additional instruction – and even more so when one gently suggests that they might continue for a whole paragraph, or more. The most famous example of a composition along these lines is Perec’s novel *La Disparition* (Perec 1969), though rather than use just one vowel, the author of this work of ludic art eliminates the letter <e> – the most common letter in French as well as in English²⁴ – in an astonishingly coherent 300-odd-page tale of the search for a missing man named Anton Voyl. Translated into languages from German to Turkish to Japanese, the likewise *e*-less English version of novelist and wordsmith Gilbert Adair (1944–2011), cleverly titled *A Void* (Perec 1994)²⁵, is well worth reading both as a virtuosic example of how to render wordplay in a foreign tongue and, indeed, in its own right.²⁶ In the words of one critic, “[i]t is almost unbelievable (yet true) that some of *La Disparition*’s first readers didn’t notice the disappearance in *The Disappearance*” (Suleiman 1990: 3).

This leads to one more question:

- (1) d. How hard is it to keep producing lipogrammatic text that looks and sounds normal enough that no one immediately notices the constraint, i.e., to produce a ludic work that does not appear to be ludic?

24 It is generally said to make up 14–15 % of any normal French text and (see, e.g., Lewand 2000: 36 and Norvig 2012) 12.49–12.70 % of any normal English one.

25 Not only is <e> avoided, but, if one allows <v> to stand in for <u>, the title contains all four other vowels in a row – A VOI – and with a void where, in a sequence that begins A, one might have expected the E. The title *A Void* (four vowels plus one consonant) is effectively the *e*-less equivalent of another title, *Eunoia* (all five vowels plus one consonant), to which I turn below in the text. By the way, Perec took all the <e>’s and used them in his 1972 novella “Les Revenentes”, whose very title – the “correct” spelling of the French noun is *revenantes* – suggests, correctly, that there will be a massive amount of “cheating”. *Vive la différence*, Derridians may say, but while Perec’s aim is to mirror the unraveling of the plot in increasingly garbled language, neither “Les Revenentes” nor the 1996 English translation of Perec’s fellow Oulipian Ian Monk (a.k.a. “E. N. Menk”), “The Exeter Text: Jewels, Secrets, Sex”, has the luminous quality of *La Disparition / A Void*.

26 An entirely *e*-less novel (or theoretically so: there are occasional slips) written originally in English is Ernest Vincent Wright’s *Gadsby* of 1939, which carries the unsubtle subtitle *A Story of Over 50,000 Words Without Using the Letter “E”*. It influenced Perec (see, e.g., Bellos 1995: 395, 399, and 402, n. 1) and has a following among wordplay fanatics; though sometimes regarded as a fine piece of literature, in my opinion it may as well be called *The Not-so-great Gadsby*. Hofstadter (1997: 103–139) offers an appreciation of *Gadsby*, *La Disparition*, *A Void*, and the art of translating lipogrammatic text and other forms of elaborate wordplay.

The answer, of course, is that it is all but impossible to manage this for more than a few sentences, especially for <u>. But there are many pleasures to be had in giving oneself over completely to alphabetic revelry. My favorite example from the current century comes from the pen of the Canadian poet Christian Bök (1966–). Directly inspired by Perec and such earlier French examples of the avant-garde as the poem “Voyelles” by Arthur Rimbaud (1854–1891) and works like *Ubu Roi* by Alfred Jarry (1873–1907), Bök wrote the brilliant *Eunoia* (Bök 2005; 1st edn. 2001), a collection whose title poem consists of five univocalic chapters, each dedicated to another experimental artist (Hans Arp, René Crevel, Dick Higgins, Yoko Ono, and Zhu Yu).²⁷ The first sentence of each chapter gives the flavor of the enterprise: *Awkward grammar appalls a craftsman; Enfettered, these sentences repress free speech; Writing is inhibiting; Loops on bold fonts now form lots of words for books*; and “*Kultur*” *spurns Ubu — thus Ubu pulls stunts*.²⁸ If it is in fact true that each vowel has a “personality”, it is hard to imagine a better way to discover it.²⁹ The titular word *eunoia* is iconic: both “the shortest word in English to contain all five vowels” and one whose meaning the poetry instantiates, for “the word quite literally means ‘beautiful thinking’” (Bök 2005: 111).³⁰ There are “many subsidiary rules” (Bök 2005: 111) aside from the obvious one, but to talk much more about Bök here – or to consider in any depth the

27 Rimbaud’s poem is the inspiration for the cover of the 2005 “upgraded edition” of *Eunoia*, and Bök prints it on p. 84 together with an English translation of his own on p. 85, as part of the section of the book titled “Oiseau”; Ubu is the protagonist of the *u*-chapter, whose first sentence is given immediately below in the text. It is worth pointing out, of Arp, who was born in Alsace and moved to Switzerland, that the *a*-chapter could not have been dedicated to Arp in his Francophone avatar, Jean. While I note, too, the different tones in the name Zhū Yù (朱昱), the infamous poem “Lion-eating poet in the stone den” by the Chinese linguist Yuen Ren Chao (e.g., https://www.youtube.com/watch?v=vExjnn_3ep4, accessed 29 July 2018) will suffice to demonstrate that tonal play in Mandarin Chinese is a fair analogue of Bök’s ludic letters.

28 The quotation marks around the first (!) word of the *u*-chapter, *Kultur*, is my way of indicating that Bök italicizes it (because it is German and not really English), unlike the rest of the sentence, which is in regular font. The relative infrequency of <u> as compared to the other vowels (see above in the text) explains why this chapter is far shorter than the other four.

29 The back cover of *Eunoia* makes the following claim: “A unique personality for each vowel soon emerges: A is courtly, E is elegiac, I is lyrical, O is jocular, U is obscene.” I find the claim annoying, both because I am not convinced that it is true and, more, because it influences readers rather than allowing them to decide for themselves whether vowels have individual personalities and, if so, what they might be. Most of my students agree that <u> has unusual potential to be obscene (compare fn. 18 above); they tend, however, to disagree, often sharply, about what the other vowels might indicate.

30 If, that is, *eunoia* (from Classical Greek εὐνοία ‘goodwill, favor’) is an English word. It has not yet made it into the *Oxford English Dictionary*.

literary qualities of either *La Disparition/A Void* or *Eunoia* – would take us too far afield.³¹ My recommendation: read the books!

4 Sounds

How should one read them? Percec's book is a novel, but Bök's is poetry and hence, one might think, possessing more than passing lyricism. At heart, though, *Eunoia* is indubitably about the alphabet, about writing, about books;³² anyone who picks up Bök's book³³ will note the special care that both author and publisher (Coach House Books in Toronto) took to make the poetry look good, from the layout of each page to the laid paper.³⁴ Nonetheless, part of the genius of *Eunoia* lies in the complex interplay between writing and sound, between how words look and what music they make. Bök, whose reading of Kurt Schwitters's "Ursonate" is available on YouTube³⁵, has made a purchasable CD-recording of *Eunoia* (Bök 2003, also from Coach House)³⁶, and one of the delights of teaching the text is getting students to read it aloud. The main

31 Still, I cannot resist pointing out one of these rules (another, especially important one is given in the following footnote): "All chapters must describe a culinary banquet, a prurient debauch, a pastoral tableau and a nautical voyage" (Bök 2005: 111). In some sense, then, *Eunoia* is a rendering of the *Odyssey*, and it should be noted – as none other than Percec (1998: 101) himself does in his "History of the lipogram" – that one Tryphiodorus around the third (?) century A.D. produced a lipogrammatic *Odyssey* (no alphas in Book 1, no betas in Book 2, etc.), a work that was explicitly derided by Joseph Addison in a 1711 essay in the *Spectator* that introduced the word *lipogram* into English (see *OED* s.v.).

32 As Bök (2005: 111) writes, "All chapters must allude to the art of writing." This is pretty clear just from the five initial sentences, reproduced above in the text.

33 Born Christian Book, he changed his name to Bök (still pronounced [bøk]) in an effort to "downplay [...] the 'embarrassment' of being an English professor and writer with the last name Book. And not just any book. 'I used to get a lot of Bible jokes when I was a kid,' he says" (Tamburri 2013). Bök's *o*-chapter contains no instances of <ö>, but forms of the word *book* are prominent on the first and last pages (Bök 2005: 59 and 76).

34 Part of the final step of Bök's short and sardonic essay "How to write Eunoia" (Bök 2002) reads as follows: "Do not bother submitting [the completed manuscript] to any publisher whose editor is not already your closest, dearest friend. Do not let the publisher do any of the design. Do it all yourself. Be a pain in the ass. Go to the publisher and commandeer a computer. Be sure take [sic] at least six weeks doing all the typesetting and copyediting".

35 <https://www.youtube.com/watch?v=VCgzRxM5010> (accessed 29 July 2018). One of the most famous instances of "sound poetry", the piece begins thus: "Fümms bö wö tää zää Uu, /pögiff, /kwii Ee."

36 At the time of writing, YouTube has a few recordings of Bök reading parts of the work.

difficulty comes from the fact that the constant repetition of a given vowel on the page sometimes but by no means always – and in a way that is entirely unpredictable – corresponds to the repetition of this vowel in sound. Take, for example, the phrase *as a papal cabal blackballs*, which appears on the first page of the *a*-chapter (Bök 2005: 12) and whose <a>'s are pronounced sequentially thus: [æ], [ə], [ei], ø³⁷, [ə], [æ], [æ], [ɔ].³⁸ In short, *Eunoia* is a type of extended tongue twister, one in which reader-cum-speakers are always in danger of losing their linguistic balance as they struggle to keep their brain and their mouth in sync.

At issue, of course, is that English spelling is an extraordinarily poor synchronic representation of the spoken language. Consider only the seven words *tough* [t^hʌf], *though* [ðou], *thought* [θɔt], *through* [θru], *thorough* [ˈθəɹiʊs] (American) or [ˈθəɹɪə] (British), *trough* [tʃɹɔf], and *thou* [ðau]: the digraph <th>, which begins five of them, is pronounced in two different ways; <ou> is pronounced in five or six different ways; and <gh> is never pronounced [g] or [h] but instead either represents [f] or is silent.³⁹ All people who know English and are literate are aware that there are precisely twenty-six letters in the alphabet, but if one asks how many sounds the language has, one receives two kinds of answers: some say twenty-six, because they confuse sounds (which are rarely taught as such) with letters (which are pedagogically fetishized from at least kindergarten), while others make wild guesses from seventy (I have almost never heard a lower number, aside from twenty-six) to 150–200 (the mode) to one or two thousand (a surprisingly common response). The correct answer depends on one's dialect, but the most common count of English phonemes is forty-four (see, e.g., Cruttenden 2014: 159 and 235). A far more interesting number than this, though, is how many representations involving the twenty-six letters there are of these forty-four sounds. By one count, the number is 1120, which means

37 The second <a> in *papal* is in some sense not pronounced (hence “Ø”): the final <l> is syllabic, [l].

38 For the angled and square brackets, see fn. 17 above; the latter are used in connection with the symbols of the International Phonetic Alphabet, a system that, at least in principle, allows anyone to sound out any sequence in any language (<https://www.internationalphoneticassociation.org/content/full-ipa-chart>, accessed 29 July 2018). Different people will have different pronunciations of some of these and many other words, which in some sense adds to the fun: in some dialects, for instance, the second <a> in *cabal* is pronounced [a]. More generally, while most will say the indefinite article *a* as [ə] in this context, [ei] is also possible.

39 A native of New York City who was not well acquainted with farms, I knew the word *trough* only from reading and was a teenager before I learned that it was not pronounced [tʃɹau]. This appears to be a common error.

that there are, on average, more than twenty-five ways of spelling each sound.⁴⁰ Yes, this includes outliers, both those that are immediately recognizable as odd – e.g., the occasional spelling *hiccough*, with <gh> = [p] – and also *women*, a common word that happens to contain the only instance of the correspondence <o> = [ɪ].⁴¹ Nonetheless, the consequences of having even a 1:10 ratio of sounds to spellings, not to mention a 1:25.5 ratio, are considerable: on the one hand, such a mismatch gives rise to spelling bees and other ludic exercises; on the other, it is a terrible burden for dyslexics, who are blessed when they grow up rather with Italian (see, e.g., Paulesu et al. 2001), in which the ratio is close to 1:1 (in fact something like 1:1.3).

In view of the discrepancy between sounds and letters, it would be surprising if the relative frequency of English sounds looked a whole lot like ETAOIN SHRDLU (or <etaoinshrdlu>). The exact order depends on dialect (and different sources report slightly different things), but a standard up-to-date reference work reports the top twelve sounds in “General British” as [əmtɪsdlrðkɛm], where [ə], [ɪ], [n], and [t] make up respectively 10.72 %, 8.29 %, 7.62 %, and 6.95 % of spoken text.⁴² Among the things to note are that the two most common sounds are vowels but eight consonants are more frequent than the third-most common vowel; that the symbol for the most common sound in the language, schwa, has no correspondent in the standard alphabet; and that there is also no correspondent in the standard alphabet for the seventh-most common consonant (3.47 % of all sounds; see Cruttenden 2014: 235), the voiced dental fricative [ð], on whose distribution see below. At the other end, things are clearer: the sound in English with the lowest functional load is the voiced palato-alveolar fricative [ʒ] (0.07 %; see Cruttenden 2014: 235)⁴³, whose rarity

40 For the figure 1120, see Helmuth (2001: 2064).

41 The word *women* is the source of the first vowel of the infamous spelling of *fish* as “*ghoti*” (on <gh> = [f], see immediately above in the text), often incorrectly attributed to G. B. Shaw but in fact predating his birth. See among much else Zimmer (2008).

42 See Cruttenden (2014: 158–159 and 235). (Cruttenden uses slashes for phonemes rather than phonetic square brackets, but the difference is immaterial here.) A mid-twentieth-century report for “General-American English” gives [əmtɪrslðdæk] (Hayden 1950: 220), where the only striking differences are (i) the near-flipping of the relative frequency of [d] and [r] (actually [ɹ]), on the second of which see Cruttenden (2014: 158); and (ii) the fact that the third-most common vowel and eleventh-most common sound in British is [e] but in American is [æ] (I note that Cruttenden 2014 makes the “substantial transcriptional alteration[]” (xvii) of changing conventional /æ/ to /a/; I have not followed him).

43 Even rarer, according to Cruttenden (2014: 159) is the diphthong [ʊə] (0.04 %), as in the Received Pronunciation of *cure*; I ignore it, however, since it is not phonemic in American English and is receding in British (see, e.g., Cruttenden 2014: 155–156).

has the evident historical explanation that it first entered English in French borrowings and is synchronically “limited to loanwords, except across word-boundaries” (Minkova 2014: 141).⁴⁴ But from the point of view of the lexicon rather than a large corpus of actual usage, the rarest sound in English is almost certainly [ð], which is found in very few discrete words, that is to say, in not more than a few handfuls of separate lemmata in a dictionary.

How can this be? Although the soothing sound that [ð] represents flows rather commonly from all speakers’ mouths, those who bother to think about the distribution of words with it should, then, find worthy of attention the way that they gather together semantically. Consider the many examples of [ð] in *this* last, admittedly awkward sentence: *the* (3x), *that* (2x), *those*, *they*, *then*, *although*, *rather*, *together*, *bother*, *gather*, *mouths*, *soothing*, *worthy*, and for some people *with*, which different people in different contexts can say as [wɪð] or *otherwise* as [wɪθ].⁴⁵ Eleven of the sixteen words (I include *with* and also, now, *this* and *otherwise*) are function words, including the two that are repeated, and there is in fact a notable paucity of content words that contain this sound. While the percentage of [ð]’s in standard discourse is quite respectable, the statistic is skewed by the extraordinary frequency of the definite article, *the*, which has the remarkable property – this is perhaps my favorite fact about English – of being by far the most common word in the language (I return to this fact in section 5) and yet consisting, in its pronunciation [ðə]⁴⁶, of nothing other than the juxtaposition of the sound that is at some level the rarest, [ð], and the one that is by no small margin the most common, [ə].

Consider now the following exercise:

- (2) a. Compose an English sentence of at least seven words in which every word has the sound [ð].

⁴⁴ Minkova (2014: 141–143) summarizes the (modest) rise of [ʒ] in English from the Middle Ages to the present. Interestingly, some speakers today have no word-initial examples of the sound at all: the only common nativized word in which I have it is *genre* (which some speakers pronounce with an initial [dʒ]), though note also *gigue* and (as Byron Ahn has impressed on me) the more recent *zhoosh*; see Minkova (2014: 142–143) for these and further examples and the speculation that “[i]t is likely that the globalisation of English will bring in more /ʒ/-initial words, as suggested by a long list of <zh>-initial words on the (unmonitored) online *Urban Dictionary*” (compare also Cruttenden 2014: 84).

⁴⁵ But not *think*!

⁴⁶ A modest curiosity of the word *the* is that it has two normal pronunciations, [ðə] and [ði]; compare the two pronunciations of the indefinite article (see fn. 38 above). For phonemic variation in common monosyllables, see Cruttenden (2014: 321).

Most wordplayers who have done this in my presence lead off with the article *the* – hardly surprising given that this is how most people begin when asked to produce a sentence about anything at all, with no constraint (compare what I wrote above in section 3 about Dirk Delabastita’s sentence). Those who have not started thus have tended instead to choose the pronoun/pronominal adjective *this* or *these* (which belong to the same dictionary lemma and are etymologically extensions of *the*), the pronoun/pronominal adjective *that* or *those* (ditto), the pronoun *they* or pronominal adjective *their*, or some other “little” word like *then*, *there*, *thus*, (*al*)*though*, *either*, or *neither*. “Correct” sentences that I have heard⁴⁷ include *The other loathsome heathens gather, breathe, bathe* and *Smothering wethers, the mothers tethered their fathers’ feathers*; neither can be called semantically normal, and the former is possible only because of the suppression of the conjunction *and* before *bathe*.

What makes the exercise interesting is precisely the fact that so few words contain the sound [ð], coupled with the tendency for these words to be grammatical building blocks. The answer to the question,

(2) b. What problems do you encounter?,

is thus: many! As it happens, the previous two paragraphs contain a not negligible percentage of the small inventory of words in the language that can be used in exercise (2a), and this means that the suggestion,

(2) c. Now try to write a follow-up sentence,

is unusually difficult to fulfill, especially if the goal is to produce something that resembles a narrative. (Still unused words include *heather*, *lathe*, *northern* and *southern*, *paths*, *rhythm*, and *thou*, though by wickedly giving just these few more I leave my readers with an even harder task of coming up with original ideas.) As for this –

(2) d. How hard is it to keep producing text that looks and sounds normal enough that no one immediately notices the constraint, i.e., to produce a ludic work that does not appear to be ludic?

⁴⁷ I stress “correct” since many people’s initial attempts have instances of [θ]: even those with linguistic training fall back under pressure on words with the voiceless counterpart of [ð], of which there are far more, though none of them is anywhere near as common (as I note briefly in section 5).

– I submit that it is impossible. But here’s a bonus exercise:

- (2) e. Compose an English sentence of any length in which every word *begins* with the sound [ð].

Readers who spend some time thinking about this should come up with both the problem and the solution, which in an effort not to spoil the fun I put into two⁴⁸ footnotes.⁴⁹

5 Words

One of the most exciting facts about language is that it obeys Zipf’s law. Noted before Harvard linguist George Kingsley Zipf (1902–1950) first popularized it in 1935 but rightly associated with him and with the mathematician Benoît Mandelbrot (1924–2010), who provided necessary modification, this power law is easy to state: the frequency of any word in a corpus of natural language (e.g., a novel, every article in the *New York Times* from 2017, the entirety of wikipédia.fr, etc.) is inversely proportional to its rank in the frequency table.⁵⁰ Put otherwise, the most common word in the language in question (ranked #1) is twice as common as the second-most common word (ranked #2), three times as common as the third-most common word (ranked #3), and so on. It does not require mathematical sophistication to spot a logarithmic curve: a very few words, like

48 The problem: the only words in English that begin with the sound are *the, this, these, that, those, there, thence, thither, they, them, their, theirs, thou, thee, thy, thine, than, then, though, and thus* – plus *themselves, thyself*, and various more or less common forms based on *there* (e.g., *thereabout(s), thereafter, thereby, therefore, therein, thereof, thereon, thereunder, thereupon, thereto, theretofore, and therewith*). Not one of them is a verb.

49 The solution: on the assumption that we disallow verbless “sentences” like *There, there!* (said in a soothing voice, with a pat on the head, to someone who is upset), the only way I know around the problem is to allow contractions: e.g., *That’s that, then!* and *They’re those*. Some will reasonably object that this is not really a solution because it stretches the definition of “word” past the breaking point.

50 See above all Zipf (1935: 39–48) but also, e.g., Zipf (1932 and 1949: 19–55 and *passim*). George A. Miller’s “Introduction” to a later printing of Zipf’s 1935 book offers a delightful look at the man and his work, which begins thus: “*The Psycho-Biology of Language* is not calculated to please every taste. Zipf was the kind of man who would take roses apart to count their petals [...]. However, for those who do not flinch to see beauty murdered in a good cause, Zipf’s scientific exertions yielded some wonderfully unexpected results to boggle the mind and tease the imagination” (Miller 1965: v).

the, are extraordinarily common; only a few hundred or a few thousand show up with any regularity at all (*family, high, help, today*, etc.); and the vast bulk of the lexicon shows up so infrequently as to be barely audible statistical background noise (*candelabra, onyx, sled, umbilical*, etc.). There is a quickly growing body of secondary literature on the Zipf-Mandelbrot law: whether it really does apply to all languages; how similar the curve for lexical frequency is to those for alphabetic and phonetic frequencies (as discussed in sections 3 and 4, respectively), which seem rather to follow the Yule-Simon distribution (see Martindale et al. 1996); what other social phenomena aside from language it applies to (e.g., the distribution of wealth (the Pareto distribution) and the population of the world's cities); how it relates to other power laws (e.g., Benford's law, which describes the frequency distribution of leading digits in numerical data sets); and – of particular interest – why it works.⁵¹ Relevant at the moment is how one can harness the statistical facts for ludic purposes.

A few years ago, computer scientist Peter Norvig, spurred on by Mark Mayzner (see fn. 19 above), made use of the Google Books NGram Viewer⁵² in order to arrive at a “distillation of the Google books data [for English that] gives us 97,565 distinct words, which were mentioned 743,842,922,321 times” (Norvig 2012). As expected, the top fifty words in English have a Zipfian distribution, with #1 *the* making up 7.14 % of the corpus, #2 *of* 4.16 %, #3 *and* 3.04 %, #4 *to* 2.60 %, etc. – down to #50 *no* 0.19 %. Here is the list:

1. <i>the</i>	11. <i>as</i>	21. <i>are</i>	31. <i>they</i>	41. <i>there</i>
2. <i>of</i>	12. <i>was</i>	22. <i>or</i>	32. <i>you</i>	42. <i>been</i>
3. <i>and</i>	13. <i>with</i>	23. <i>his</i>	33. <i>were</i>	43. <i>if</i>
4. <i>to</i>	14. <i>be</i>	24. <i>from</i>	34. <i>their</i>	44. <i>more</i>
5. <i>in</i>	15. <i>by</i>	25. <i>at</i>	35. <i>one</i>	45. <i>when</i>
6. <i>a</i>	16. <i>on</i>	26. <i>which</i>	36. <i>all</i>	46. <i>will</i>
7. <i>is</i>	17. <i>not</i>	27. <i>but</i>	37. <i>we</i>	47. <i>would</i>
8. <i>that</i>	18. <i>he</i>	28. <i>have</i>	38. <i>can</i>	48. <i>who</i>
9. <i>for</i>	19. <i>I</i>	29. <i>an</i>	39. <i>her</i>	49. <i>so</i>
10. <i>it</i>	20. <i>this</i>	30. <i>had</i>	40. <i>has</i>	50. <i>no</i>

⁵¹ On the Zipf-Mandelbrot law, see recently Piantadosi (2015), Sorrell (2015), Williams et al. (2015), and Lestrade (2017). For a quirky application of the methodology of complex adaptive systems to the language (and culture) of furniture (!), see Burkette (2015: 55–57 and *passim*).

⁵² <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> (accessed 29 July 2018).

These fifty vocabulary items account for about 40.5 % of the material in a given text!⁵³ Furthermore, the list has all sorts of properties that are for the most part as obvious as they are remarkable: for example, every single word is monosyllabic (and none has more than five letters)⁵⁴ and, indeed, a proverbial “Anglo-Saxon monosyllable” since the only ones that are not quite native are *they* and *their*, which are shallow early borrowings into English from the closely related North Germanic languages; every single word is a function word (whether article, conjunction, preposition, pronoun, copula, basic (always potentially auxiliary) verb – plus a few other forms like *all*, *more*, *no*, and *not*); seven of the words have (or can have) the sound [ð] (*the*, *this*, *that*, *they*, *their*, *there*, and sometimes *with*; see section 4), while its voiceless counterpart, which is found in far more individual words of, however, far lower functional load (*think*, *thorax*, *pathological*, *myth*, etc.), is absent; and the least-common vowel, *u*, is

53 Note that Norvig’s list relies on the idea that one word is distinguished from another purely through external form. It would take us too far afield to consider in detail the vexed question of what exactly the definition is of the term “word” (see now Wray 2015), but among the things to think about are: whether *he* and *his* count as different words, and the same for *they* and *their*; whether *a* and *an* count as different words; whether *is*, *are*, *was*, *were*, *be*, and *been* count as six different words (or whether the pair *be/been* is one word and/or the synchronically connected but irregular and yet diachronically explicable pair *was/were* and/or the synchronically connected but irregular and yet diachronically unconnected (!) pair *is/are*), and the same for *have* and *had* and even, perhaps, for *will* and *would*; whether it is reasonable to lump all instances of *to* together, and the same for some other forms, such as *for* and *that*; and whether we should care that some instances of *can* in the corpus are containers rather than modal verbs and some instances of *will* a noun with one or another meaning. Three contributions to *The Oxford Companion of the Word* (Taylor (ed.) 2015) aside from Wray (2015) are of particular relevance to this paper: Sorrell (2015) on “Word frequencies”, Grzybek (2015) on “Word length”, and Verkuyl (2015) on “Word puzzles”.

Many scholars work with the data found at <https://www.wordfrequency.info> (accessed 29 July 2018), which is based on the 450 million-word Corpus of Contemporary American English. Some differences with Norvig’s list are clear just from the top twenty: 1. *the*, 2. *be*, 3. *and*, 4. *of*, 5. *a*, 6. *in*, 7. *to*, 8. *have*, 9. *to*, 10. *it*, 11. *I*, 12. *that*, 13. *for*, 14. *you*, 15. *he*, 16. *with*, 17. *on*, 18. *do*, 19. *say*, and 20. *this*. Here all synchronically allied forms of *be* are lumped together, as are all forms of *have*, and *a* includes also *an*; by contrast, two meanings of *to* are distinguished, as are two of *that* (the other one is #27). The appearance of the verbs *do* and *say* is interesting (as are *n’t* at #29 (compare fn. 49 above); the North Germanic loanword *get* at #39 (on *they* and *their*, see immediately below in the text); and a lone disyllable, *about*, at #46 – historically a prefixed form of obsolete *bout* and a word whose first syllable can be omitted in colloquial speech (*’bout*)).

54 The relationship between brevity of form and frequency of use is well known: see, e.g., Zipf (1935: 20–39) and Grzybek (2015: 107, with references).

found only three times (*but, would, and you*)⁵⁵, with just the one having it as its only vowel (and thereby usable by Bök in *Eunoia*; see section 3).⁵⁶

There is great potential here for play:

- (3) a. Compose an English sentence of at least seven words that does not use any of the ten most-frequent words in the language.

No one finds this difficult: examples that have come my way include *Because bananas smell sweet, rascally monkeys love them* and *She sells seashells by seaweedy, sushi-laden seashores*.⁵⁷ But because this is so easy, the matter of the coherent follow-up is, I believe, especially interesting:

- (3) b. Now try to write a follow-up sentence, perhaps one that uses none of the twenty most-frequent words. (Then try to write a third sentence without using any of the thirty most-frequent words. Etc.)

Readers are encouraged to try their hand at this exercise and to pay attention to the point at which it starts to become really tough – and why this is. As always, the question looms:

- (3) c. How hard is it to keep producing text that looks and sounds normal enough that no one immediately notices the constraint, i.e., to produce a ludic work that does not appear to be ludic?

The most astonishing creative work I know that wreaks havoc on the fact that words are not distributed at all evenly is Seattle-based Doug Nufer's novel *Never Again* (Nufer 2004). Two hundred pages and about 40,000 words long, it has a perfectly level lexical distribution since no word – by which is meant set of characters set off by a space on either side – is ever repeated (!). The novel's first sentence is perfectly ordinary: *When the racetrack closed forever I had to get a job*. Of these eleven words, six are in the top fifty – *the* (#1), *to* (#4), *a* (#6), *I* (#19), *had* (#30), and *when* (#45) – and only one, *racetrack*, is not especially common. Not one of these words can be used ever again. Here is the first paragraph in its entirety (Nufer 2004: 3):

55 By contrast, there are eighteen *e*'s, fourteen *o*'s, thirteen *a*'s, and eleven *i*'s.

56 Another thing to note: the pronouns *I, you, he, it, we, and they*, all of which are (or can be) conventionally used as subjects, are all more common than *she*, which is not in the top fifty (in fact, it is #51); the conventional objective form *her*, however, is there, though ranked below all the others.

57 Readers who grew up with English will recognize this as a variation on the well-known tongue twister *She sells seashells by the seashore*, which contains the illicit word *the*.

When the racetrack closed forever I had to get a job. Want ads made wonderlands, founding systems barely imagined. Adventure's imperative ruled nothing could repeat. Redirections dictated rigorously, freely. Go anywhere new: telephone boiler-rooms, midnight grocery shooting galleries, prosthetic limb assembly plants, hazardous waste-removal sites; flower delivery, flour milling, million-dollar bunko schemes. Do anything once; then, best of all, never again.

Two other words in the top fifty show up here – *of* (#2) and *all* (#36) – and the circumstances render thrilling the rhetorical interplay between the ordinary yet iconic (*Do anything once*) and the wacky (*prosthetic limb assembly plants*). In short, the relationship between content and form is palpable, as is Nufer's evident delight in low-level wordplay (*Want ads made wonderlands* and *flower delivery, flour milling, million-dollar*). And how does the work close? Here is the last paragraph (Nufer 2004: 202):

Worldly bookmaker soulmates rectify unfair circumstance's recurred tragedies, ever-moving, ever-hedging shifty playabilities since chances say someone will be for ever closing racetracks.

In addition to making a ring between initial *When the racetrack closed forever* and final *for ever closing racetracks*, playing with compounded forms of *ever*-, and bringing his linguistic train to an iconic halt, Nufer unveils in his final seven-word clause three of the top fifty words, which, astonishingly, he has saved until the very end: *for* (#9), *be* (#14), and *will* (#46) – plus, just before this, the very common verb *say* (see fn. 53 above) and the conjunction *since*. Truly a virtuoso work in the Oulipian spirit!

6 The future: After Scrabble

Humans are creative, and we can be confident that the generations ahead will produce new and unexpected forms of wordplay. There is, however, one aspect of play with words that I expect to see actively highlighted in the coming years, and it involves paying close attention to what exactly speakers and readers of English count as a word. Every January, the members of the American Dialect Society vote on the “Word of the Year” for the previous twelve months. In 2017, the winner was *fake news*; in 2016 it was *dumpster fire* (a poor choice, in my

view); and in 2015 it was singular *they*.⁵⁸ We may ask whether *fake news* and *dumpster fire* are actually each an individual word, but a more interesting question concerns the Word of the Year for 2014: #blacklivesmatter. Are this and #metoo really “words”? How do we know? Do they belong in a dictionary? If so, how are they to be alphabetized?

And yet even these questions pale, I believe, next to the challenge to the idea of the word posed by the team at the Oxford Dictionaries, which at the end of each calendar year announces *its* Word of the Year. In 2017, the winner was *youthquake* (another poor choice); in 2016 it was *post-truth*; and in 2014 it was *vape*. But what about in 2015? In that year, the people who might be said to hold the greatest authority over the English lexicon informed the world that the Word of the Year was ... 😊 (“face with tears of joy”).⁵⁹ The rise of emoji is socio-culturally fascinating for reasons that go well beyond the ludic, but in a brave new world in which 😊 may count as a word and in which it is possible to purchase *Emoji Dick*, a “crowd sourced and crowd funded translation” of *Moby-Dick* (<http://www.emojidick.com>, accessed 29 July 2018), future students in “Wordplay: A wry plod from Babel to Scrabble” and similar classes will have much to think about and much to look forward to.

In his definition of “play” quoted in fn. 8 above, Huizinga claims that wordplay is “an activity connected with no material interest, and no profit can be gained by it”. There is truth to this, but at the same time a significant benefit of actively engaging in verbal games is that one gains an appreciation of just what language can do, of how far it can be bent without snapping. Furthermore, actively engaging with and pushing against the boundaries of the ball of language – of paying attention, e.g., to what Dr. Seuss’s *One Fish, Two Fish, Red Fish, Blue Fish* and James Joyce’s *Finnegans Wake* do and do not have in common – is, I contend, an outstanding way of getting students interested in linguistics, a subject that vanishingly few encounter even in high school, never mind earlier still. Percec has said that constraints offer freedom, and to this I would add another paradox: marginal phenomena shed significant light on the core.

Acknowledgment: My thanks go to Esme Winter-Froemel for suggesting that I lead a workshop on the teaching of wordplay at the conference in Trier whose

⁵⁸ See <https://www.americandialect.org/fake-news-is-2017-american-dialect-society-word-of-the-year> and <https://www.americandialect.org/woty/all-of-the-words-of-the-year-1990-to-present> (both accessed 29 July 2018).

⁵⁹ See <https://en.oxforddictionaries.com/word-of-the-year> (accessed 29 July 2018).

proceedings are collected here and to the many participants, especially Dirk Delabastita, Ray Gibbs, Cécile Poix, and Eline Zenner. I first read “The Vane sisters” in the summer of 1986 under the brilliant tutelage of Lawrence Raab and the late Lawrence Graver, whose Telluride Association Summer Program seminar “Art and mystery” at Williams College was life-changing. I also thank my own students, above all the sixty once-freshmen whose enthusiastic contributions to my seminar on wordplay at Princeton have taught me so much and brought such joy. I dedicate this paper to “the Larrys” and all my fellow TASPers and to the members of the 2009 inaugural class at Princeton, one of whom, the ever-laughing Emilly Zhu ’13, whose photograph and one of whose specific contributions appear in Noden (2010), died on 12 June 2015 after having been hit by a car.

7 References

- Adams, P. G. & T. Cable. 2012. Alliteration. In Roland Greene (ed.), *The Princeton Encyclopedia of Poetry and Poetics*, 4th edn., 40–42. Princeton: Princeton University Press.
- Augarde, Tony. 2011. *Wordplay: The Weird and Wonderful World of Words*. Charlbury: Carpenter.
- Bellos, David. 1995. *Georges Perec: A Life in Words*, rev. edn. London: Harvill.
- Bök, Christian. 2002. How to write Eunoia. *Broken Pencil* 19. 72.⁶⁰
- Bök, Christian. 2003. *Eunoia*, CD version. Toronto: Coach House.
- Bök, Christian. 2005. *Eunoia*, “upgraded edn.” from 2001 original. Toronto: Coach House.
- Borgmann, Dmitri A. 1965. *Language on Vacation: An Olio of Orthographical Oddities*. New York: Scribner’s.
- Burkette, Allison Paige. 2015. *Language and Material Culture*. Amsterdam: Benjamins.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Conradt, Stacy. 2011. 22 fascinating and bizarre courses offered this semester. *Mental Floss* (28 August 2011). <http://mentalfloss.com/article/28626/22-fascinating-and-bizarre-classes-offered-semester> (accessed 29 July 2018).
- Cook, Guy. 2000. *Language Play, Language Learning*. Oxford: Oxford University Press.
- Cruttenden, Alan. 2014. *Gimson’s Pronunciation of English*, 8th edn. Abingdon: Routledge.
- Crystal, David. 1998a. From Scrabble to babble: Reflections on language attitudes and language play. In Wolfgang Kühlwein (ed.), *Language as Structure and Language as Process: In Honour of Gerhard Nickel on the Occasion of His 70th Birthday*, 33–45. Trier: Wissenschaftlicher Verlag.
- Crystal, David. 1998b. *Language Play*. London: Penguin.
- Crystal, David. 2008. *Txtng: The Gr8 Db8*. Oxford: Oxford University Press.

⁶⁰ The author of Bök (2002) is mistakenly named “Bok”, which is amusing in view of the material in fn. 33 above.

- Eckler, Ross. 1996. *Making the Alphabet Dance: Recreational Wordplay*. New York: St. Martin's.
- Erard, Michael. 2010. The life and times of "Colorless green ideas sleep furiously". *Southwest Review* 95(3). 418–425.
- Gordin, Michael D. & Joshua T. Katz. Forthcoming. The walker and the wake: Analysis of non-intrinsic philological isolates. In Sean Gurd & Vincent W. J. van Gerven Oei (eds.), *'Pata-philology: An Irreader*. N.p.: punctum.
- Grzybek, Peter. 2015. Word length. In Taylor (2015), 89–119.
- Hayden, Rebecca E. 1950. The relative frequency of phonemes in General-American English. *Word* 6(3). 217–223.
- Helmuth, Laura. 2001. Dyslexia: Same brains, different languages. *Science* 291, no. 5511 (16 March 2001). 2064–2065.
- Hofstadter, Douglas R. 1985. *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Basic Books.
- Hofstadter, Douglas R. 1997. *Le Ton beau de Marot: In Praise of the Music of Language*. New York: Basic Books.
- "Hottest college courses". 2011. *The Daily Beast* (6 October 2011). <http://www.thedailybeast.com/galleries/2011/10/06/hot-college-courses-mad-men-south-park-and-more-unique-classes.html> (accessed 29 July 2018).
- Huizinga, J. 1949. *Homo ludens: A Study of the Play-element in Culture*. London: Routledge & Kegan Paul.
- Jahn, Manfred. 2002. "Colorless green ideas sleep furiously": A linguistic test case and its appropriations. In Marion Gymnich, Ansgar Nünning & Vera Nünning (eds.), *Literature and Linguistics: Approaches, Models, and Applications. Studies in Honour of Jon Erickson*, 47–60. Trier: Wissenschaftlicher Verlag.
- Katz, Joshua T. 2007 [publ. 2008]. An acrostic ant road in *Aeneid* 4. *Materiali e discussioni per l'analisi dei testi classici* 59. 77–86.
- Katz, Joshua T. 2008. Vergil translates Aratus: *Phaenomena* 1–2 and *Georgics* 1.1–2. *Materiali e discussioni per l'analisi dei testi classici* 60. 105–123.
- Katz, Joshua T. 2009. Wordplay. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 20th Annual UCLA Indo-European Conference, Los Angeles, October 31–November 1, 2008*, 79–114. Bremen: Hempen.
- Katz, Joshua T. 2010a. Etymology. In Anthony Grafton, Glenn W. Most & Salvatore Settis (eds.), *The Classical Tradition*, 342–345. Cambridge, MA: Harvard University Press.
- Katz, Joshua T. 2010b. *Nonne lexica etymologica multiplicanda sunt?*. In Christopher Stray (ed.), *Classical Dictionaries: Past, Present and Future*, 25–48. London: Duckworth.
- Katz, Joshua T. 2013a. Gods and vowels. In J. Virgilio García & Angel Ruiz (eds.), *Poetic Language and Religion in Greece and Rome*, 2–28. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Katz, Joshua T. 2013b. The Muse at play: An introduction. In Jan Kwapisz, David Petrain & Mikołaj Szymański (eds.), *The Muse at Play: Riddles and Wordplay in Greek and Latin Poetry*, 1–30. Berlin: De Gruyter.
- Katz, Joshua T. 2013c. Saussure's *anaphonie*: Sounds asunder. In Shane Butler & Alex Purves (eds.), *Synaesthesia and the Ancient Senses*, 167–184. Durham: Acumen.
- Katz, Joshua T. 2014a. Acrostic. In Richard F. Thomas & Jan M. Ziolkowski (eds.), *The Virgil Encyclopedia*, [1.] 8. Malden, MA: Wiley-Blackwell.
- Katz, Joshua T. 2014b. Puns. In Georgios K. Giannakis (ed.), *Encyclopedia of Ancient Greek Language and Linguistics*, 3. 193–194. Leiden: Brill.

- Katz, Joshua T. 2014c. Riddles. In Georgios K. Giannakis (ed.), *Encyclopedia of Ancient Greek Language and Linguistics*, 3. 244–245. Leiden: Brill.
- Katz, Joshua T. 2014d. Wordplay. In Richard F. Thomas & Jan M. Ziolkowski (eds.), *The Virgil Encyclopedia*, [3.] 1396–1397. Malden, MA: Wiley-Blackwell.
- Katz, Joshua T. 2015. Saussure at play and his structuralist and post-structuralist interpreters. *Cahiers Ferdinand de Saussure* 68. 113–32.
- Katz, Joshua T. 2016a. Another Vergilian signature in the *Georgics*?. In Phillip Mitsis & Ioannis Ziogas (eds.), *Wordplay and Powerplay in Latin Poetry*, 69–85. Berlin: De Gruyter.
- Katz, Joshua T. 2016b. Etymological ‘alterity’: Depths and heights. In Shane Butler (ed.), *Deep Classics: Rethinking Classical Reception*, 107–126. London: Bloomsbury.
- Knospe, Sebastian, Alexander Onysko & Maik Goth (eds.). 2016. *Crossing Languages to Play with Words: Multidisciplinary Perspectives* (The Dynamics of Wordplay 3). Berlin & Boston: De Gruyter.
- Lawson, JonArno. 2008. *A Voweller’s Bestiary from Aardvark to Guineafowl (and H)*. Erin, ON: Porcupine’s Quill.
- Lestrade, Sander. 2017. Unzipping Zipf’s law. *PlosOne* (9 August 2017). <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181987> (accessed 29 July 2018).
- Lewand, Robert Edward. 2000. *Cryptological Mathematics*. Washington, D.C.: Mathematical Association of America.
- Levin Becker, Daniel. 2012. *Many Subtle Channels: In Praise of Potential Literature*. Cambridge, MA: Harvard University Press.
- Martindale, Colin, S. M. Gusein-Zade, Dean McKenzie & Mark Yu. Borodovsky. 1996. Comparison of equations describing the ranked frequency distributions of graphemes and phonemes. *Journal of Quantitative Linguistics* 3(2). 106–112.
- Mayzner, M. S. & M. E. Tresselt. 1965. *Tables of Single-letter and Digram Frequency Counts for Various Word-length and Letter-position Combinations. Psychonomic Monograph Supplements* 1(2) = pp. 13–32.
- Miller, George A. 1965. Introduction. In George Kingsley Zipf, *The Psycho-biology of Language: An Introduction to Dynamic Philology* [new edn. of Zipf (1935)], v–x. Cambridge, MA: MIT Press.
- Minkova, Donka. 2014. *A Historical Phonology of English*. Edinburgh: Edinburgh University Press.
- Muldoon, Paul. 2001. *Poems, 1968–1998*. New York: Farrar, Straus & Giroux.
- Noden, Merrell. 2010. Word freaks. *Princeton Alumni Weekly* 110(8) (3 February 2010). 24–26.
- Noegel, Scott. 2007. *Nocturnal Ciphers: The Allusive Language of Dreams in the Ancient Near East*. New Haven: American Oriental Society.
- Norvig, Peter. 2012. English letter frequency counts: Mayzner revisited, or ETAOIN SRHLDCU. <http://norvig.com/mayzner.html> (accessed 29 July 2018).
- Nufer, Doug. 2004. *Never Again*. New York: Black Square.
- Paulesu, E., J. F. Démonet, F. Fazio, E. McCrory, V. Chanoine, N. Brunswick, S. F. Cappa, G. Cossu, M. Habib, C. D. Frith & U. Frith. 2001. Dyslexia: Cultural diversity and biological unity. *Science* 291, no. 5511 (16 March 2001). 2165–2167.
- Perec, Georges. 1969. *La Disparition*. Paris: Denoël.
- Perec, Georges. 1977. “Des Règles pour être libre” [interview with Claude Bonnefoy]. *Les Nouvelles littéraires* 2575 (10 March 1977). 21.
- Perec, Georges. 1994. *A Void*, trans. of Perec (1969) by Gilbert Adair. London: Harvill.

- Perec, Georges. [1969] 1998. History of the lipogram. In Warren F. Motte, Jr. (trans. and ed.), *Oulipo: A Primer of Potential Literature*, rev. edn., 97–108. Normal, IL: Dalkey Archive.
- Piantadosi, Steven T. 2015. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review* 21(5). 1112–1130.
- Alloway, Lee C. F. 1985. In quest of a pangram. *Abacus* 2(3). 22–40.
- Sorrell, Joseph. 2015. Word frequencies. In Taylor (2015), 68–88.
- Stephens, Greg J. & William Bialek. 2010. Statistical mechanics of letters in words. *Physical Review E* 81, no. 066119.
- Stump, Jordan. 2003. Exercises in wile: Raymond Queneau, the novelist as trickster. *Bookforum* 10(3). 12–14.
- Suleiman, Susan Rubin. 1990. *Subversive Intent: Gender, Politics, and the Avant-garde*. Cambridge, MA: Harvard University Press.
- Tamburri, Rosanna. 2013. The incredibly original pursuits of Christian Bök. *University Affairs / Affaires universitaires* (6 November 2013). <https://www.universityaffairs.ca/features/feature-article/incredibly-original-pursuits-of-christian-bok/> (accessed 29 July 2018).
- Taylor, John R. (ed.). 2015. *The Oxford Handbook of the Word*. Oxford: Oxford University Press.
- Thaler, Verena. 2016. Varieties of wordplay. In Knospe, Onysko & Goth (2016), 47–62.
- Verkuyl, Henk J. 2015. Word puzzles. In Taylor (2015), 702–722.
- Waxman, Olivia B. 2015. 11 bizarre college courses we actually want to take. *Time* (26 August 2015). <http://time.com/4006878/unusual-college-university-courses/> (accessed 29 July 2018).
- Williams, Jake Ryland, Paul R. Lessard, Suma Desu, Eric M. Clark, James P. Bagrow, Christopher M. Danforth & Peter Sheridan Dodds. 2015. Zipf's law holds for phrases, not words. *Scientific Reports* 5, no. 12209.
- Winter-Froemel, Esme. 2016. Approaching wordplay. In Knospe, Onysko & Goth (2016), 11–46.
- Winter-Froemel, Esme & Angelika Zirker (eds.). 2015. *Enjeux du jeu de mots. Perspectives linguistiques et littéraires* (The Dynamics of Wordplay 2). Berlin & Boston: De Gruyter.
- "Wordplay winners". 2010. *Princeton Alumni Weekly* 110(11) (7 April 2010). 30–31.
- Wray, Alison. 2015. Why are we so sure we know what a word is? In Taylor (2015), 725–750.
- Zimmer, Ben. 2008. "Ghoti" before Shaw. *Language Log* (23 April 2008). <http://languagelog.ldc.upenn.edu/nll/?p=81> (accessed 29 July 2018).
- Zipf, George Kingsley. 1932. *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press.
- Zipf, George Kingsley. 1935. *The Psycho-biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton Mifflin.
- Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.
- Zirker, Angelika & Esme Winter-Froemel (eds.). 2015. *Wordplay and Metalinguistic/Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-reflection* (The Dynamics of Wordplay 1). Berlin & Boston: De Gruyter.

