David Crystal

Still Finicking with Trifles, Largely
(The Otto Jespersen Memorial Lecture, 1995)

This title calls out for deconstruction. The context, as you would expect, given this occasion, is a remark made by Jespersen in his farewell lecture on retiring from Copenhagen University in 1925:

> To anyone who finds that linguistic study is a worthless finicking with trifles, I would reply that life consists of little things; the important matter is to see them largely.

It is a dictum promoted by some textlinguists to begin an analysis with the lexical item which carries the highest information content in a sentence, namely *trifle*, in the present instance; but I have little to say about this, other than to point out that in my title it does not refer to what the *Longman Dictionary of English Language and Culture* defines as 'a British dish made of plain cakes set in fruit and jelly covered with cream and/or CUSTARD' (the small caps identifying not emphasis, but a cross-reference to an entry in which *custard* is glossed as 'especially British English' - itself a not insignificant trifle, which previously I am not sure I had been aware of). Rather, it is sense 2, glossed as formal, 'an article or thing of little value or slight importance; matter of slight importance'.

Finicking is more fascinating. An unusual use, as a verb, my intuition says. *I finick? You finick? They finicked? John and Mary were finicking in the lane?* The Longman Dictionary does not give *finick* as a verb, but only as an adjective, *finicky*, glossing it as 'too concerned about unimportant details', or 'fussy'. Its bigger stable-companion, the *Longman Dictionary of the English Language*, does include *finicking*, but as an adjective, not as a verb, and also has an adjectival use of *finical*, but without the *k*, glossed simply as 'finicky'. Let us change stables. There is nothing additional in my Chambers desk dictionary. Nor in my Collins. Nor in my Concise Oxford. Nor in Partridge's *Dictionary of Slang*. I start to panick.

Fortunately, the big unabridged dictionaries are helpful. The OED, for example, has adjectival *finicky* and also *finicking*, the latter with an alternative spelling *finikin*, and a verb *finick*, found only in a dialectal British use, with various spellings, and illustrated by a single 19th-century quotation which shows *finnicking* with two *n*'s (three if you count the one in the -*ing*). The noun *finick* is said to be †*Obs* (obsolete). The OED Supplement adds verb examples, including citations from George Meredith and Virginia Woolf, suggesting that the usage was becoming standard - and having found examples of the noun as late as 1949, tells us soberly to 'Delete †*Obs*'.

Now all kinds of linguisticky things come to mind when one reflects on *finick*. The collocations, for example: what, lexically, goes with *finick*? *Trifles*, according to Jespersen, and I have no problem with that. But what else? Indeed, *is* there anything else? Or is *finick*

with trifles an idiom? Not according to the OED examples, where a wide range of possible contexts is suggested.

> The verse laughs at such finnicking [...]
> The Demon pointed his feet... and finicked a few steps away.
> That sort of person finicks with the marriage question.
> I do not finick about fearing what people think [...]

Webster III adds:

> [...] finicked with her food (Elizabeth Taylor)
> she was not one who had time to finick about snipping at blossoms

These examples suggest that the verb has a very wide collocational range, but what are the constraints? There must be constraints. Lexical items do not collocate with everything. Prepositionally, we can evidently *finick* both *with* and *about,* or neither. But what may we finick with/about? May we say *They finicked with their friends? He finicked with his correspondence? I was finicking with my soup?* I do not know. Nor do I think do you.

Let me now take a more familiar word, where mutual ignorance about collocational norms seems to operate. *Dictum.* What do we do to a dictum? If I stumble across a dictum of, say, Jespersen, and report it to you using the construction 'This dictum, Verbed by Jespersen', which verb should I use? Do we *coin* a dictum, or *formulate* one, or *present* one, or *announce* one? *Made, given,* and *used* seem tame or not quite right. *Propounded, pronounced,* and *promulgated* are all possible, but somewhat ponderous. *Delivered, voiced, advanced, introduced, adumbrated,* and several other verbs are all available, but all add special nuances. Is there no basic, neutral, standard collocation in English?

Dictionaries exist to provide remedies for failed intuitions. Unfortunately I could find no example of a transitive verb governing *dictum.* The OED provided only an instance of *adduce* in a legal context. An informant test on half-a-dozen people brought no consensus - only more verbs (*mooted, framed, exclaimed* [...]). It is difficult to resist the conclusion that people do not know what to do to a dictum, and that even some quite unusual collocations might pass without comment. For example, can we *promote* a dictum? A dictum is 'a formal statement of opinion' (Longman): *promote,* in the most applicable sense, is 'to help in the growth or development of' (as in 'promoting world peace'). On the face of it, there would seem to be no need to help develop something which has achieved its final formulation. But did anyone notice, when I said, 1000 words ago, a grief ago indeed, just after my first Jespersen quotation, "It is a dictum promoted by some textlinguists [...]"?

These are examples of what I believe to be a hidden iceberg of lexical fuzziness in English. I do not know how many lexical items would betray such indeterminacy, upon investigation, but I am sure there are thousands. Moreover, such examples are important to help provide a corrective to the simplified accounts of collocation which often appear in textbooks, where *spick* convincingly collocates with *span, quench* with *thirst,* and *auspicious* with *occasion.* I have unassailable intuitions about all of these. But I do not know what you do to a dictum, or what you can finick with.

In case you were wondering, let me say at this point what *I* think my lecture title is about. It is, as I have suggested, about fuzziness, not fussiness - fuzziness in relation to both our intuitive knowledge of our language and in relation to the boundaries of our academic knowledge. I am concerned that our textbooks and teaching materials present a picture of the

English language (or, for that matter, any language) which is altogether too determinate, and, by using such labels as 'comprehensive' or 'unabridged' in their titles, give the impression that most of the relevant information is known. If we restrict ourselves to the study of grammar, we might be justified in claiming that, at least for the written language, what we know far exceeds what we do not know. But when we consider the lexicon, the opposite is the case. This is not, in fact, an area which Jespersen dealt with very systematically. His works, such as *Mankind, Nation and Individual*, of course provide a fascinating collection of lexical anecdotes, and he was continually drawing attention to the general linguistic principles which influence our study of words; but whatever his general contribution to lexical study might be said to be, it did not warrant a separate section in the retrospective volume by Juul and Nielsen, for example, and Jespersen's own fascination with grammar, combined with the recency of semantics as a systematic approach, led him to apologise for not investigating the lexical aspects of his subject in the *Philosophy of Grammar* (1924:35).

Things have not greatly changed, since Jespersen's time. The lexicon has continued to receive short shrift from linguists (as opposed to lexicographers and language teachers) in recent years - a peculiar irony, when one realizes that, although grammar is the foundation or skeleton of a language (select the metaphor of your choice), most of its 'weight' or 'flesh' consists of its vocabulary (at least a million words, as we shall see, in the case of English). Also, this is the area where we get closest to meaning, which is what we are all chiefly concerned with. I am reminded of a remark of Jespersen's at one point in his career, while explaining a change in his intellectual direction: "I wanted something pleasurable to do and thought syntax more attractive than morphology". Now there's a professional. For most people, whatever the words *pleasurable* and *attractive* collocate with, it is unlikely to be either syntax or morphology. But his reason is interesting - because, he says, syntax gets to grips with meaning more directly. And that is my point, too, in focusing on the lexicon.

Finick, for me, is a symbol of the indeterminacy and ignorance which surrounds lexical study. *Finick*. The very word is like a bell [...]. But of course, it isn't like a bell. Not a real, macho bell, anyway. Keats's *forlorn* is, indeed, bell-like - a serious bell, as the contemporary idiom would put it. But *finick* is as far away from the vowel symbolism of that kind of bell as you can get - a pair of weedy short high front vowels, compared with the muscular long back open vowels of *forlorn*. Or, if you want another example, the resonating long open diphthongs of "the tolling bell, Measures time not our time, rung by the unhurried Ground swell [...]" (Eliot, "The Dry Salvages"). That also sounds like a serious bell. If *finick* were to be bell-like, it would have to be of the tinkling type, a little, titchy, wee, twee, teeny-weeny sort of bell.

We are, it would appear, into sound symbolism - onomatopoeia, if you prefer, though the issues involved go well beyond the literary associations conveyed by that term. This is one of the least-investigated aspects of the lexicon - and for reasons which are well known, at least in linguistics. Sound symbolism has traditionally had something of a bad press, due to the rigid distinction between sounds and meanings insisted upon by dualistic linguistic theories, and the number of counter-examples it is always possible to find when one attempts to establish a correlation between a particular type of sound and a particular type of meaning. Sounds are not supposed to have meanings, and this has been one of the reasons for the difficulty linguists have had in explaining our phonaesthetic intuitions about literature. On the other hand, the evidence is slowly growing that the outright rejection of sound symbolism has been premature. Several studies in phonetics and psycholinguistics are drawing our attention to the existence of phonetic (as opposed to phonological) patterns which can help to explain variation in usage. The correlations are more subtle than the crude correspondences postulated

by early accounts, such as the proposal that vowel height related to object size - a proposal supported by my *bell* example, as it happens, but famously called into question by such counter-examples as *big*, which should be tiny, and *small*, which should be large.

This point was brought home to me by a 1990 study which helped to explain my phonaesthetic intuitions about two phonetic effects I had been worrying about during the late 1980s. In 1986, the satirical UK TV programme *Spitting Image* recorded "The Chicken Song", in which the lyrics invited the listener to perform a range of bizarre activities, such as (as I recall) bury all your clothes, paint your left knee green, climb inside a dog, and (the climax of the first verse) pretend your name is Keith. Why is it bizarre to be 'Keith'? A couple of years later, British comedian Rowan Atkinson, as Captain Blackadder, in a First World War trench, encounters a pretty girl dressed as a male soldier. Wanting to keep her for himself, and not wishing to give away her identity to his colleagues, he gives her a male name: 'Bob' - to the delight of the audience, who then laugh each time he uses the name. Why is 'Bob' funny?

In 1990, an interesting phonological analysis of the structure of English first names was published in the *Journal of Linguistics* (by Anne Cutler and two colleagues at Cambridge). They analysed nearly 1700 items from a dictionary of first names, looking at the differences between male and female names. This is what they found. (I omit the statistical support, and suggest you use your intuitions to test the claims they make.)

- Female first names tended to be longer than males, in terms of the number of syllables they contain. Males are much more likely to have a monosyllabic first name (*Jim, Fred, John*), and much less likely to have a name of three or more syllables (*Christopher, Nicholas*). By contrast, there are few monosyllabic female first names (*Ann, Joan, May*) and many of them are trisyllabic or more (*Katharine, Elizabeth, Amanda*).

- 95% of male names have a first syllable which is strongly stressed, whereas only 75% of female names show this pattern. It is not difficult to think of female names which begin with an unstressed syllable (*Patricia, Elizabeth, Rebecca*), but male names are very rare (*Jerome, Demetrius*). In fact, none of the popular British names in the frequency lists in the last 75 years has had an unstressed initial syllable.

- The stressed syllables of female names tend to make much more use of the high front vowel /iː/, such as *Lisa, Tina, Celia, Maxine*, and the archetypal *Fifi* and *Mimi*. Male names in /iː/ are far less common - *Peter, Steve, Keith*.

- Female pet names tend to be longer than male. A bisyllabic pet name could be either male or female, but a monosyllabic one is much more likely to be male. *Jackie* could be either sex, but *Jack* is male. (Other examples include *Bob/Bobbie* and *Bill/Billie*.)

- Female names are much more likely to end in a (spoken) vowel, as with *Linda, Tracey, Patricia, Mary*. If not a vowel, the last sound will very likely be a continuant, especially a nasal, as in *Jean, Kathleen, Sharon, Ann*. By contrast, plosives are much more likely to be found in male endings (*David, Dick, Jock*).

Interesting questions arise. For example, is *Kate* more male-sounding than *Kath* or *Katie* or *Katherine*? When Henry V of England meets Princess Katharine of France (*Henry V*, 5.ii), he calls her both *Katherine* and *Kate*. But he uses *Katherine* only with a preceding attribute - usually *fair*, but also *dear*, *la plus belle*, and *Queen of all* - each a strongly female collocation. When Henry uses a straightforward vocative, it is *Kate* - an appropriate pet form, perhaps,

from a 'plain king' who knows 'no ways to mince it in love', and who speaks to her as a 'plain soldier'.

These are observations, not explanations. Is there some basis for the sound symbolism? Can such associations as smallness and brightness, often linked with the /i:/ vowel, explain the preference for /i:/ in the female names? Can we relate the trend towards the use of an initial stressed syllable to greater masculine aggressiveness? Certainly, if I were a script writer, and I had to think up the most inappropriate name for a girl dressed as a man, the above tendencies would lead me to choose a monosyllabic form, using a closed syllable, ending in a consonant as far away from a continuant as I can find - a plosive - and with a vowel as far away from /i:/ as I can find, such as /a/ or /ɒ/. *Bob*, in short. (You can speculate about *Keith* for yourselves, and - to balance the examples - I leave you also to consider why, in a recent US survey, a sample of American men overwhelmingly judged the sexiest female name to be *Christine*.)

We may move this example, haltingly, in the direction of poetry. In a *Sunday Times* poll of British readers' favourite words in 1980, *melody* and *velvet* tied for first place. Third was a tie between *gossamer* and *crystal*, followed by *autumn*, *peace*, *tranquil*, *twilight*, and *murmur*, with *caress*, *mellifluous*, and *whisper* tying for tenth place. The occasion seems to have motivated a poem by John Kitching.

I like to think of words with lovely sounds
That I can ease around my Sunday tongue—
—Like velvet, melody and young,
Gossamer, crystal, autumn, peace,
Mellifluous, whisper, tranquil, lace,
Caress and silken, willow, mellow,
Lullaby, dawn and shimmer, yellow,
Silver, marigold and golden,
Dream and harmony and olden,
Blossom, champagne, sleep and dusk,
Magic, hummock, love and mist,
Darling, laughter, butterfly,
Charity, eiderdown and sky,
And parakeet and rosemary,
Froth, gazebo, ivory,
And syllabub and vacillate,
Mesmerism, echo, fate,
Jacaranda, harlequin
And chrysalis and violin,
Enigma, tart and sycamore,
Pomp, chinchilla, truffle, myrrh,
Bewildered, claret, akimbo, fur,
Flamingo next and celandine,
Ominous, tantalise and wine,
Antimacassar, jewel, skill,
Russet, buckram, delight and thrill,
Clavichord and didgeridoo,
Doppelganger, fractious, zoo.
I don't know what they mean. Do you?
But I like to have them in my head
And dandle them and handle them
Like Wedgwood china. What finer?
(John Kitching, "Sunday Words": 1980)

What is it about the phonaesthetics of these words which makes them so attractive? Which vowels and consonants are most involved? If there were time, it would be a useful exercise to stop at this point, and jot down the sounds which strike you as particularly important, before comparing your list with the results of a systematic phonetic survey. The task, of course, is to notice not only which sounds are frequently used, but also—rather more difficult—which are not used at all.

An analysis of the 81 words listed in the poem shows some clear trends.

- The consonants divide into two types: high frequency and low frequency. Just eight items account for 73 per cent of all consonants (264): /l/ has 41 instances (15 per cent), followed by /m/ (27), /s/ (25), /k/ (23), /r/ (21), /t/ and /d/ (19), and /n/ (18). If this ranking is compared with that found in conversation, the use of /l/ and /m/ is noteworthy.
- There is then a big jump before reaching the low-frequency consonants: /f/ and /b/ (9), /p/ and /v/ (8), /g/ (7), /z/ (6), /ŋ/ (5), /w/ (4), /ʃ/, /tʃ/ and /h/ (3), /θ/, /dʒ/ and /j/ (2). Only /ð/ and /ʒ/ do not occur at all.
- If we group these consonants into types according to their manner of articulation, frictionless continuants are commonest (118:68 oral, 50 nasal), followed by plosives (85), fricatives (56), and affricates (5). As there are only four oral continuants (/l, r, w, j/) and three nasals (/m, n, ŋ/), but six plosives (/p, b, t, d, k, g/) and nine fricatives (/f, v, θ, ð, s, z, ʃ, ʒ, h/), this distribution is noteworthy.
- Of the 172 vowels, the unstressed vowel /ə/ is commonest (43), showing that words of more than one syllable are preferred. Only 21 words were monosyllables; the largest category (28) was words of three syllables. Most were stressed on the first syllable; and most made use of at least three different manners of consonant articulation.
- The other common vowel was / ɪ / (38), which occurred over twice as often as the next vowel /a/ (16), followed by /əʊ/ (11), /e/ (10), /i:/, /ʌ/, and /aɪ/ (9), /ɒ/ (6), /ɑ:/ (5), /eɪ/, /u:/ and /ɔ:/ (4), /ɜ:/ (2), /aʊ/ and /aɪə/ (1). This is close to the vowel rankings of conversation.

Is John Kitching's intuition representative of *Sunday Times* readers? Of the 68 vowels and consonants used in the paper's words, only 13 are missing from his top eight consonants and top eight vowels. There is an 80 per cent chance that the readers would like his other words too.

This analysis perhaps explains why a serious romantic poem about London Underground stations would very likely include *Pimlico* and *Colindale*, which closely reflect these intuitions, and exclude *Goodge Street* and *Wapping*, which do not. Also why friendly space aliens receive such names as *Alaree* and *Osnomian*, why enemy names include *Vatch* and *Triops*, and why *Klingons* are likely to be a mite less aggressive than *Kryptons*. Also why, if we wanted to create a new word which was phonaesthetically correct, it would seem advisable to give it three syllables, stress the first, use at least one /m/ and /l/ (preferably both), vary the manner of articulation, and keep most vowels short. We would probably find success with *ramelon* and *drematol.* On the other hand, we could simply settle for "The moan of doves in immemorial elms / And murmur of innumerable bees" (Tennyson, "The Princess").

It may be of interest to note that, if *finick* were a first name, on the above basis it would be of ambivalent sex, the initial stress and final -*k* suggesting masculinity, the high front vowels suggesting the opposite. However, *finick* is not a first name, and the only reason for that

sentence is to bring us back, in the manner of the link so favoured by news reporters ("They may be playing cricket in South Africa, but in this Lancashire village cricket couldn't be further from their minds [...]") to my continuing explication of the title of this paper, where we move on now to a different aspect of the basic indeterminacy and ignorance which, I am suggesting, characterizes our study of the lexicon: how do you spell *finick*? Earlier references show it with and without a *k*, and with either one or two *n*'s. Evidently there is some uncertainty. Is this a trifling matter?

How do you spell *finick*? We are brought up in a literacy tradition which insists on a definite answer to such a question. We expect there to be a single correct spelling for any word in the standard language; and if we do not know what it is, we expect to find an unambiguous answer in a dictionary. However, the reality is somewhat more complex.

To begin with, there are several cases where people have no idea how to spell a word - where it seems impossible to spell a word in an acceptable way, whatever you do. One study (Abbott, 1988) collected examples of words with unusual endings, and asked how an *-ed* or *-ing* ending might be added to them.

a(h) polka, verandah, visa, mascara, umbrella, samba, sauna, aroma, balaclava, tiara
e(e) purée, flambé, recce (reconnaissance), frisbee, tree
et parquet, bouquet, beret, duvet, chalet, ballet
i ski, sari, jacuzzi, bikini

The problem is evident. What is the past tense of *samba*? Does it look right to put *They sambaed* or *We're sambaing*? Some write *samba'd* or *samba-ing*. Professional writers vary in their decisions: David Lodge has "her heavily mascaraed eyelids" (*Small World*, 1984:125); Frederick Forsyth has "So get visa-ed up in Paris" (*The Dogs of War*, 1974:117). Dictionaries are usually silent on this sort of point. Indeed, I know several famous writers who have complained about being unable to find an awkward spelling in a dictionary; on the other hand, point this out to the lexicographers, and they simply say that they are waiting for examples from these very writers to give them some data on which to base their recommendations. (I am reminded of Thomas Hardy's dislike of critics, who often condemned his neologisms. Robert Graves, in *Good-bye to All That* (1929) tells a nice story: "Once or twice recently he [Hardy] had looked a word up in the dictionary for fear of being again accused of coining, and found it there right enough - only to read on and discover that the sole authority quoted was himself in a half-forgotten novel!")

Sometimes the addition of an ending produces a conflict of readings (though context makes real ambiguity unlikely): *they skied* (from *ski* or *sky*); *an anoraked figure* (where the spelling suggests a long vowel pronunciation /eɪ/ for the second *a*); *the current arced* (where pronouncing *c* as /s/ before *e* does not hold [sic], leading some writers to prefer *arcked*).

The author of this study concludes by giving a short paragraph containing some of his own preferences. I doubt whether any two educated English writers would arrive at an identical set of solutions.

I would rather be in a comfortable verandahed house, sitting pyjamaed in a duveted bed and being fed puréed fruit by a muumuued beauty, than be bivouacked on a sparsely-treed plain, sitting anorak-ed and shivering in the leaden-skied gloom and eating potatoes that were sautéd yesterday before the power cables arc-ed.

This leads us on to the remarkable number of alternative spellings in Standard English. Some are well known, such as the differences between British and American English (e.g. *-our/-or*),

the use of optional *e* (*judg(e)ment*), and the choice between *-ise/-ize*. Another area is the question of whether a word should be hyphenated or not (as in compounds: *flowerpot, flower-pot*, or *flower pot*) or spelled with a capital letter (*babel* or *Babel*).

The pages of an unabridged dictionary will bring to light many examples. Here is the result of a cull of just one page of entries: the beginning of letter *B* in *Webster's Third New International Dictionary*. The symbol * indicates that the alternatives may also be capitalized.

```
baa / ba
baal-ha-bos / balabos
baalshem / balshem
babacoote / babakoto
babasco / barbasco
babassu oil / babaçu oil
babaylan / babailan /
babalyan / babalian
babbitt / babbit*
babbittry / babbitry*
babes-ernst / babes-ernest
babirusa / babirousa / babirussa
babu / baboo
babul / babool
```

There were in addition 19 cases where the only difference between the words was the use of a capital letter: *baal* (*-ism, -istic, -ite*), *babbittical, babbitty, babcock test, babel* (*-ism, -ization, -ize*), *babi, babinski reflex, babism, babist, babouvism, babouvist, babylon, babylonian*. Including these cases, this page had 32 items with alternative spelling out of a total of 95 entries – a third. Excluding them, we are still left with 14 entries (a remarkable 15 per cent).

Was my page atypical? Sidney Greenbaum carried out a similar survey (1986), analysing all the spelling variants for letter A in two British desk dictionaries. He excluded all names with specific reference (proper names), counted just once any items which belonged to more than one part of speech (*ascendant/-ent* as noun or adjective), and did not count separately items listed with irregular inflectional endings (*acouchis/-ies*), but did include lexical derivations (*antagonize, antagonizable*, etc.). He found 296 entries with variable spellings, averaging 3 a page, and calculates that the proportion of entries in the work as a whole is 5.6%. Only 21% of these, incidentally, are due to the difference between British and American systems. He concludes:

> Spelling variation is found at different levels in the vocabulary of British English. We encounter it in the informal *aunty, auntie*; in the ordinary everyday *aging, ageing*; in the poetic *aery* (in place of *airy*); in the learned *anomie, anomy*; and in the technical *arytenoid, arytaenoid*. There is probably greater uniformity in spelling than in any other aspect of the English language, and nonstandard spellings are widely regarded as a sure mark of the uneducated. Yet the evidence of this paper shows that the spelling of standard English is by no means almost invariant [he is referring here to a quotation from a standard text, which claimed that it was so]. And the variance is most conspicuous in the learned and technical vocabulary, which we might expect to display greatest uniformity and of which only a small fraction appears in general dictionaries [...].

That is why my sample page contains so much more variation, for this sample was taken from an unabridged dictionary, the bulk of which comprises technical or exotic terms. Indeed, exotic loan words present the English spelling system with the biggest modern challenge to its consistency. And as English continues its inexorable growth around the world, encountering

more languages in the process (over 120 now) and increasingly borrowing from them as localities try to impose their own identity on linguistic colonial tradition, we must expect this problem to grow. On the other hand, the growing use of spelling checkers and related technological pressures will undoubtedly act as a strong standardizing constraint. So the future remains unclear.

It is extraordinary that so many of the obvious questions to do with the lexicon remain unanswered. Elementary questions, in a way, such as the overall size of the lexicon, or the size of some subset of it. Questions of size are among the most frequently asked questions by the general public. How many words does a 5-year-old know or use? How many words are there in Shakespeare, or in anyone? How many words are there in the language, anyway? Linguists are well aware of the problems. There are different answers to the question, depending on what counts as a word. Do you include proper names, abbreviations? Do multi-word idioms count as a single lexical item? Do items such as *go*, *going*, and *gone* count as one? Depending on your answer to such questions, your count will vary greatly. Still, even allowing for these methodological considerations, the totals are conspicuous by their absence (literary concordance work being the honourable exception). We know how many words Shakespeare used, but not how many a 5-year-old child uses.

All we can be sure about is that the totals cited in received wisdom are a long way from the truth, and that the revisions are consistently upwards. We see this in the primitive child language studies, where earlier reports of vocabulary levels in 3-year-olds numbering several hundred, maybe a thousand, have to be replaced by several thousand (Wagner 1985). Most of these low estimates arose from the way the data were obtained - counting the vocabulary in children's readers, for example, and assuming that these would bear some correspondence to spoken reality. Bridie Raban's report (1988) on the spoken vocabulary of 5-year-old children killed such stereotypes once and for all. Her report lists such items as *kill*, *gun*, *bombs*, and *bullets* – presumably (I hope) reflecting a world of play – as well as *bum*, *bugger*, *titties*, and *fart* (this last as frequent as *bath* and *clock*, and, interestingly, as *God* and *Guinness*). There is something fascinating about any survey in which *Henny Penny* turns up in the same frequency range as *machine gun*. It seems there can be murder even in fairyland.

When we turn our attention to the question of the size of the English lexicon as a whole, the inadequacy of our present accomplishments becomes even more apparent. The two biggest dictionaries each suggest around half a million lexical items – a total approached by the unabridged *Webster's Third New International* (which claimed over 450,000 entries in 1961) and by the integrated edition of the *Oxford English Dictionary* (which claimed over 500,000 entries in 1992). The true figure is undoubtedly a great deal higher.

A comparison of these two dictionaries – or of any other group of dictionaries of comparable size–shows a remarkable lack of identity between headword lists. In one sample, from *saba* to *sabbaticalness*, analysed for the *Cambridge Encyclopedia of the English Language* (p. 119), the *Webster* and *Oxford* have only 21 headwords in common out of a possible 57 – less than two-fifths. If this pattern were continued, their combined lexicon would exceed three-quarters of a million. And this is a headword study: we have not begun to look at the correspondence between senses yet.

Discrepancies are usually caused by differing editorial emphases. The *Oxford* has far more historical references and British dialect items than does the *Webster*, which in turn has far more local American items. On the other hand, neither work would claim to be comprehensive in its coverage of the vocabulary of the "new Englishes" in such parts of the world as India, Singapore, and Nigeria, where thousands of new lexemes are coming into the

language. And because the tradition in lexicography is to use the written language as the test for inclusion, much local spoken nonstandard vocabulary will be omitted. There must be thousands of slang expressions currently in common use which have never been recorded, such as all the lexemes which express the concept of 'being drunk'.

Even if we restrict the issue to standard vocabulary, there are many items which could be included as part of the lexicon, but which are not usually found in a dictionary. There are some half a million abbreviated forms in English, many of which have a clear lexical status; and fauna and flora also provide a vast lexical resource. For example, there are apparently some million insects already described, largely in Latin, with several million more awaiting description. This means that there must be at least a million loan-word designations enabling English-speaking entomologists to talk about their subject. Should all of these be allowed into the word-count as well?

It is difficult to see how even a conservative estimate of English vocabulary could go much below a million lexemes. More radical accounts, allowing in all of scientific nomenclature, could easily double this figure. Aware of these issues, over a decade ago Lawrence Urdang was one who was arguing for a superdictionary project. If we extend the Webster/OED comparison to include other dictionaries, you begin to get a sense of the limitations of the term 'unabridged'. The combined total of the two dictionaries, you will recall, was 57. Reference to *Chambers English Dictionary* (a much shorter work) brings to light another five items. Reference to Willis's *Dictionary of the Flowering Plants and Ferns* (8th edn) gives another six. We have reached over 70 items now, with many other specialist dictionaries within the UK left to consult, and the varieties of English in the world at large breathing heavily in the wings. My comment, some time ago, that the English vocabulary has at least a million words, must surely be a gross underestimate.

This is finicking on the largest scale. We have left the trifle metaphor a long way behind. It is, I hope, the kind of finicking of which Jespersen would have approved, for it is finicking with a large purpose. The purpose is easily stated: to put the lexicon once again in centre stage, where it used to be, and to devote to it the same kind of intellectual energy that we have devoted to grammar and phonology. It is for this reason that, in the sections on structure, in the *Cambridge Encyclopedia of the English Language*, I have devoted twice as much space to the lexicon than to grammar. And why is it important to treat the lexicon with respect? Because it is the chief way of meeting what Jespersen in his farewell lecture called a first requirement in historical linguistic research: "an understanding of the texts as a matter of pure philology in the narrow traditional meaning of the word: to penetrate into the innermost thoughts of the best men and women." That dictum, with the omission of "best", I would enthusiastically promote today, and especially at this conference, with its wide-ranging subject-matter. For it is in the area of the lexicon, not in grammar and phonology, that I believe we will find the most fruitful way forward of relating our various concerns - linguistic and literary, historical and modern - and thus for giving our subject of English studies a unifying focus. We can finick, I fink, no more largely than that.

References

Abbot, G. (1988): "Mascaraed and muumuu-ed: the spelling of imported words". In: English Today 14, 43-46.
Crystal, D. (1995): The Cambridge Encyclopedia of the English Language. Cambridge: C.U.P.
Cutler, A./J. McQueen/K. Robinson, (1990): "Elizabeth and John: sound patterns of men's and women's names." In: Journal of Linguistics 26, 471-482.

Greenbaum, S. (1986): "Spelling variants in British English." In: Studies in English Linguistics, 19, no. 2, 258-268.

Jespersen, O. (1924): The Philosophy of Grammar. London: Allen & Unwin.

– (1946). Mankind, Nation, and Individual. London: Allen & Unwin.

Juul, A./H.F. Nielsen (1989): Otto Jespersen: Facets of his Life and Work. (Amsterdam: John Benjamins) (= Studies in the History of the Language Sciences 52).

Raban, B. (1988): The Spoken Vocabulary of 5-year-old Children. Reading: Reading and Language Information Centre.

Summers, D. (ed.) (1992): Longman Dictionary of English Language and Culture. Harlow: Longman.

Wagner, K.R. (1985): "How much do children say in a day?" In: Journal of Child Language 12, 475-487.