English Orthography

The one-l llama,
He's a priest.
The two-l llama,
He's a beast.
And I will bet
A silk pajama
There isn't any
Three-l-llama.

("The Lama," Ogden Nash)

The one-l llama is a Tibetan word that appeared on European lists in the mid-17th century. The two-l llama is a Spanish word (borrowed from Quechua) that appeared on European lists in the early 17th century. The three-l llama was invented by Ogden Nash in the 1930s and it currently appears on no European lists. All three are spelled correctly above.

Orthography (aka, spelling) is complicated in English because our words come from so many sources. Professional writers seem to agree that one has to understand spelling and grammar in order to play with it as Nash does. Most of us memorize or look up some words we write, although “sounding-out” was the method taught when I was in school. Unfortunately, it’s not entirely reliable.

Dan Quayle did more to make people aware of spelling difficulties than any recent public figure with his “p-o-t-a-t-o-e.” Ironically, according to Ammon Shea, “there have been no fewer than sixty-four different ways of spelling potato over the past 450 years, and that’s counting only the ones that begin with P...” (A neighbor when I was a child spelled it “budado,” which was exactly how she said it.) Lexicographer Shea adds, “almost all of these are regionalisms or archaic spellings... but there is one that has had surprising longevity, appearing in print almost throughout almost all of the twentieth century—the humble potatoe.” Although the San Francisco Chronicle, the Salt Lake City Tribune and the Washington Post used that spelling during the last two decades of the twentieth century, the standard spelling was (supposedly) p-o-t-a-t-o and still is. (David Crystal perhaps explains Quayle’s problem when he points out that “most words ending in an -o preceded by a consonant have their plurals spelled with
-oes." Bur, of course, there are several common exceptions, such as albinos, pianos, banjos, and [Yikes!] potatos.)

Attempts to spell phonetically can lead to disaster (not dis-assed-her). Here are a few supposedly unedited samples from Twitter:

"So hot in the office one of the secretaries thought she was metal pausing."

"Twitter is my alter eogie."

"I love it when you can still smell your girlfriend's colon on you."

"I always ovary act to small stuff."

"Time heals all wombs...I hope that's true."

"My girlfriend and I went from being together all the time to barley talking."

Articulation and hearing clearly play a part in creating those gems, and so does typing too fast and not proof-reading.

Every language has a corpus of phonemes—sounds that are recognized as unique and used contrastively (thus "fat-gat-cat-rat-mat," etc., are different because their initial sounds contrast). In some spelling systems, signs are assigned to each sound (or combination of sounds). One source of spelling problems for us is that most English dialects employ 44 distinct sounds (phonemes) but we have only 26 letters with which to represent them. As a result, a letter combination such as -ough is used to represent different sounds in the words "tough, though, thought, thorough, through, bough, cough," and "hiccough."

Crystal points out that "a widely cited estimate" is that English spelling is about 75% regular, but that many of the 400 or so most common irregular words are among the most frequently used. As a result, some sounds or combinations of sounds seem ubiquitous. For example, which words in the following sentence seem to you to have "regular" and "irregular" spellings? "Did he believe that Caesar could see the people seize the seas?" Problems may arise because we have so many ways to write common sounds such as /i/ (long ee) in the sample sentence.

A quick look at the complicated history of English explains why it can be difficult for spellers. The language's development started when various Celtic tribes crossed onto Britain from mainland Europe in about 500 BC (remnants of their speech survives in some English place names (such as Avon, Dover, Cam, Dee, etc.) and a few other English words like bin, crag, curse, and dun. The Romans followed (43 BCE-410 AD) and they brought a different language structure and a much richer culture and vocabulary (some early borrowings include cheese, kettle, cup, wine, dish, mountain, gate, etc.). The complicated process of English-language development was then profoundly influenced by Germanic tribes (Angles, Saxons, Frisians and later Jutes) who invaded and created the structural base for Old English (contributing even affixes and "structure words" such as un-, for, with, mis-, of, on, etc.). In the
sixth century the so-called Anglo-Saxons encountered Christian missionaries, who contributed from Latin some religious terms primarily (abbot, nun, angel, disciple, etc.). Priests also got spelling off to a bad start when they tried to squeeze the 35 (or thereabouts) distinct sounds of Old English into their 23-letter Latin alphabet.

During this time and for the several centuries to follow Scandinavian raiders were part of British experiences (bringing words like scatter, want, skill, cut, bask, etc.). In 1066 matters came to a head when two Viking chiefs—one who by then thought of himself as English and another who thought of himself as Norman—fought for supremacy of Normandy and England (the Norman won) and brought a rush of French vocabulary and spelling to the language (employed at first mostly by the upper classes), far too many to list but virtually the whole vocabulary of law, of government, of art, and of pleasure.

"By the beginning of the fifteenth century," reports Crystal, "English spelling was a mixture of two systems—Old English and French." During the sixteenth century and beyond Latin and Greek words were adopted, sometimes self-consciously ("inkhorn terms"), and the English vocabulary was much expanded although spelling was again disrupted (Latin additions include: executor, simile, prosecute, explicit, item, etc.; Greek additions include: anonymous, lexicon, tonic, idiosyncrasy, criterion, etc.). A strange phenomenon called the Great Vowel Shift occurred between roughly 1400 and 1600 when the pronunciation of some vowel sounds systematically shifted. The results of the shift can be heard in the modern pronunciation of the altered vowels in words like please-pleasant, serene-serenity, sane-sancty, crime-criminal, etc. Their spelling remained unchanged while their pronunciation was different.

Two more big influences have blown through English spelling, scattering letters and sensibilities: the growth and influence of American English; language historian Lincoln Barnett describes it as "the weedlike growth of the American vernacular as it spread across prairies, deserts, and mountain to the Pacific Ocean, thirstily absorbing foreign words and phrases on every frontier...." (and creating orthographic contrasts like colour-color, musick-music, theatre-theater, defence-defense, travelling-traveling, etc.). More recently, the explosion of various electronic media and their vocabularies (TV, email, tweet, ebook, weblog, etc.)—also largely blamed on Americans—continues growing and mutating. Recognizing that spelling's purpose has not only been to identify words, but to defend heritage...or so it seems...it is difficult to overestimate the outrage those developments have caused traditionalists.

Perhaps as a result, the inconsistency of the English spelling system has long been fodder for humorists. A verse by Richard Krogh points out some problems in English spelling: "Beware of heard, a dreadful word/ That looks like beard and sounds like bird./And dead: it's said like bed, not beard;/For goodness sake don't call it deed!/Watch out for meat and great and threat./They rhyme with suite and straight and debt./A moth is not a moth in mother./Nor both in bother, broth in brother."

What follows is a list of the 30 most commonly written English words according to the Oxford English Corpus of over a billion words ("including writing of all sorts from literary novels to specialized journals to everyday newspapers and magazines and from Hansards [Parliamentary transcripts] to the language of chatrooms, emails and weblogs," [from Wikipedia]):
1. the; 2. be; 3. to; 4. of; 5. and; 6. a; 7. in; 8. that; 9. have; 10. it; 11. it; 12. for; 13. not; 14. on; 15. with; 16. he; 17. as; 18. you; 19. do; 20. at; 21. this; 22. but; 23. his; 24. by; 25. from; 26. they; 27. we; 28. say; 29. her; 30. she.

The first 25 of those words make up about one-third of all printed material in English, according to The Reading Teachers Book of Lists, while a list of the first 100 (see Wikipedia) will account for fully half...and all of them are easy for most people to spell.

The 20 most commonly misspelled (Brit. misspelt) English words, according to Oxford University Press (2112) are separate, definitely, maneuver (Brit. manoeuvre), embarrass, occurrence, consensus, unnecessary, acceptable, broccoli, referred, bureaucracy, supersede, questionnaire, connoisseur, a lot (alot), entrepreneur, particularly, liquefy, conscience, parallel. Just missing that list, ironically enough, is the word misspell (mispell).

Robert N. Feinstein has perhaps summarized matters, tongue in cheek, in his poem “Gnormal Pspelling”:

Gnus and gnomes and gnats and such-
Gnouns with just one G too much...
Pseudonym and psychedelic—
P becomes a surplus relic.
Knit and knock an knife and knocked—
Kneedless Ks are overstocked.
Rhubarb, rhetoric and rhyme
Should lose an H from thyme to time.