

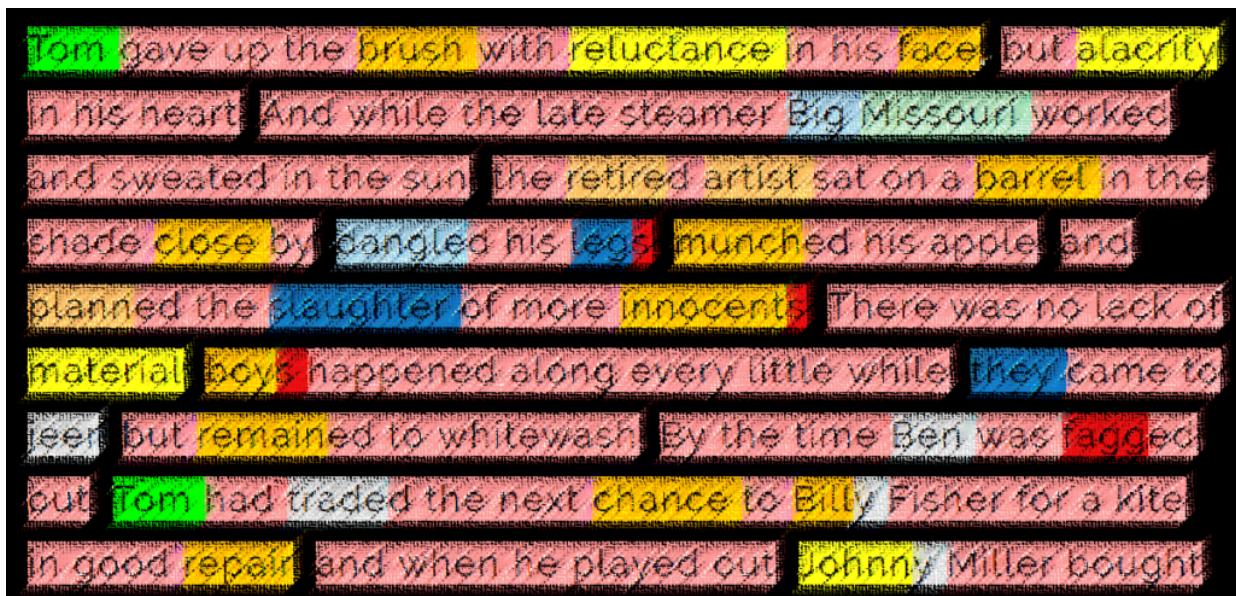
Ideas Illustrated

Survival Skills for the Information Age

Menu

[About](#) [Resources](#) [Subscribers](#)

Menu



Tom gave up the brush with reluctance in his face, but alacrity in his heart. And while the late steamer *Big Missouri* worked and sweated in the sun, the retired artist sat on a barrel in the shade close by, dangled his legs, munched his apple, and planned the slaughter of more innocents. There was no lack of material: boys happened along every little while, they came to jeer, but remained to whitewash. By the time Ben was fagged out, Tom had traded the next chance to Billy Fisher for a kite in good repair, and when he played out, Johnny Miller bought

Visualizing English Word Origins

Posted on April 1, 2012

I have been reading a book on the development of the English language recently and I've become fascinated with the idea of word [etymology](#) — the study of words and their origins. It's no secret that English is a great borrower of foreign words but I'm not enough of an expert to really understand what that means for my day-to-day use of the language. Simply reading about word history didn't help me, so I decided that I really needed to see some examples.

Using Douglas Harper's online [dictionary of etymology](#), I paired up words from various passages I found online with entries in the dictionary. For each word, I pulled out the first listed language of origin and then re-constructed the text with

some additional HTML infrastructure. The HTML would allow me to associate each word (or word fragment) with a color, title, and hyperlink to a definition.

The results look like this:

The quick brown fox jumps over the lazy dog.

This simple sentence is constructed of eight distinct words and one word suffix. Six of the words are from Old English (colored in pink) while the others are from Gallo-Roman and Middle Low German (both colored in gray). Hovering over each word provides the exact source and clicking the word takes you to the full origin description.

A second example shows more variety:

Supreme executive power derives from a mandate from the masses, not from some farcical aquatic ceremony.

This is a surprisingly complex Monty Python quote where the colors represent Old English (pink), Middle English (red), Anglo-French (orange), Old French (light orange), Middle French (pale orange), and Classical and Medieval Latin (both yellow). I suspect that both the complexity and variety of word sources is intentional — standing in humorous contrast to the appearance of the [speaker](#).

What follows are five excerpts taken from a spectrum of written sources. The intent was to investigate each passage and see if word origin varied significantly based on the intended purpose of the passage.

(This process was pretty involved and my initial dream of creating an app that would allow me to convert any paragraph to this format faded when I realized that much of the word matching process needed manual intervention. I definitely suggest digging in to the full etymology site to explore the full history of each word. I have probably made plenty of translation mistakes as I developed my paragraphs but I certainly had fun.)

Passage #1: American Literature

The first paragraph I looked at was an excerpt from Mark Twain's *The Adventures of Tom Sawyer*. I chose this text because I thought it would have a good mix of

English and American words.

Tom gave up the brush with reluctance in his face, but alacrity in his heart. And while the late steamer Big Missouri worked and sweated in the sun, the retired artist sat on a barrel in the shade close by, dangled his legs, munched his apple, and planned the slaughter of more innocents. There was no lack of material; boys happened along every little while; they came to jeer, but remained to whitewash. By the time Ben was fagged out, Tom had traded the next chance to Billy Fisher for a kite, in good repair; and when he played out, Johnny Miller bought in for a dead rat and a string to swing it with — and so on, and so on, hour after hour. And when the middle of the afternoon came, from being a poor poverty-stricken boy in the morning, Tom was literally rolling in wealth. He had beside the things before mentioned, twelve marbles, part of a jews-harp, a piece of blue bottle-glass to look through, a spool cannon, a key that wouldn't unlock anything, a fragment of chalk, a glass stopper of a decanter, a tin soldier, a couple of tadpoles, six fire-crackers, a kitten with only one eye, a brass door-knob, a dog-collar — but no dog — the handle of a knife, four pieces of orange-peel, and a dilapidated old window sash .

The passage has a solid base of Old English words mixed with a variety of French, Latin and Old Norse terms. Middle English makes an appearance in the form of a few words and suffixes while American English is found solely in the list of items Tom Sawyer collects from his friends. Two of these American terms ("fire-crackers" and "door-knob") are hyphenated words built from Old English and Scandinavian components. (Several of Twain's other hyphenated words apparently didn't make it over the hump into full-fledged Americanisms. However, it should be noted that Twain was often the first author to record usage of U.S. slang of the era.)

I found it interesting that Middle English had such a poor showing in this text but it may be due to the fact that the defining elements of Middle English have more to do with sentence structure and grammatical elements than specific words. I was also surprised at the frequent use of longer, Latin-based words in an adventure novel, but the average word length comes in at about 4.4 characters — still fairly short and simple.

Although 73% of the word fragments are Old English, Twain uses words from over a dozen different sources in this short passage alone. Overall, the wide variety of

word sources adds a pleasing “flavor” to the passage. The mix seems well-balanced and interesting.

Passage #2: British Literature

For my second test, I wanted to look at text from a non-American author. I chose a paragraph from Charles Dickens' ~~A Tale of Two Cities~~ *Great Expectations* out of respect for my 7th-grade English teacher.

My sister had a trenchant way of cutting our bread-and-butter for us, that never varied. First, with her left hand she jammed the loaf hard and fast against her bib —where it sometimes got a pin into it, and sometimes a needle, which we afterwards got into our mouths. Then she took some butter (not too much) on a knife and spread it on the loaf, in an apothecary kind of way as if she were making a plaster —using both sides of the knife with a slapping dexterity, and trimming and moulding the butter off round the crust. Then, she gave the knife a final smart wipe on the edge of the plaster, and then sawed a very thick round off the loaf: which she finally, before separating from the loaf, hewed into two halves, of which Joe got one, and I the other.

The relative simplicity of this passage surprised me a little. The average word length is about 4.2 and over 84% of the word fragments are basic Old English. No other source comes in over 5% and the variety of sources is half that of the Twain passage. ~~American English~~ Hebrew makes an appearance in the form of the name “Joe” but most of the other borrowed words are French in origin. Still, I found the text appealing in a way — basic words for a basic task.

Passage #3: Legal

The third paragraph comes from a United Nations document on maritime territories. I selected this passage because it seemed to contain more jargon and I suspected that much of this jargon was borrowed. This hunch proved to be correct.

Where the coasts of two States are opposite or adjacent to each other, neither of the two States is entitled, failing agreement between them to the contrary, to extend its territorial sea beyond the median line every point of which is equidistant from the nearest points on the baselines from which the

breadth of the territorial seas of each of the two States is measured. The above provision does not apply, however, where it is necessary by reason of historic title or other special circumstances to delimit the territorial seas of the two States in a way which is at variance therewith.

This text had a much higher ratio of French and Latin word fragments (16.9% and 9.3%) and a longer average word length — nearly 4.8 characters — than both previous passages. With 64.4% of the word fragments, Old English still serves as a major binding agent in this text but there is less variety overall. Middle English makes its appearance only as a suffix and there is only one word outside of the English/French/Latin triumvirate. After the visual and poetic excitement of the two literature entries, this paragraph seems very bland.

Passage #4: Medicine

Note: This passage has been revised (see thread)

My dad suggested that I take a look at a healthcare-related passage to see if the use of specific medical terminology would tilt the word usage even farther away from “native” English words. Boy, was he right.

The anatomic axis of the lower extremity is defined by the femorotibial angle, which averages 5° of valgus; the mechanical axis of the lower extremity is defined by a plumb line connecting the center of the femoral head to the mid ankle on a standing anteroposterior weight-bearing radiograph. The mechanical axis averages 1. 2° of varus, and it is more accurate than the anatomic axis in demonstrating load transmission across the knee joint, especially if femoral or tibial deformities contribute to limb malalignment. A study by Khan et al in patients with early symptomatic knee osteoarthritis showed a clear relationship between local knee alignment (as determined from short fluoroscopically guided standing anteroposterior knee radiographs) and the compartmental pattern and severity of knee osteoarthritis. In this study, each degree of increase in the local varus angle was associated with a significantly increased risk of having predominantly medial compartment osteoarthritis, and a similar association was found between the valgus angulation and lateral compartment osteoarthritis in 47 knees.

The medical paragraph has only 51.9% Old English word fragments and the average number of characters per word is 5.7 — much higher than even the legal

text. French Latin, and Greek were used more frequently in this passage and, despite U.S. prowess in the healthcare field, there were no American English terms. This is a paragraph that is doing a lot of heavy lifting and it uses a lot of dense, muscular words to get the job done.

Passage #5: Sports

This last passage was an attempt to stack the deck in favor of some home grown words. It doesn't get more American than baseball, but the only American word in this [article](#) about a spring training rainout between the Milwaukee Brewers and the Texas Rangers is the word "baseball" itself. Everything else is either Old English or borrowed. Still, I have to assume that phrases like "at-bats" and "suicide squeeze bunt" are not exactly common constructions and my guess is that the entire article would be a mystery to someone who didn't know the game.

It was a wild, windy day at Maryvale Baseball Park before the rains came with the Brewers ahead, 6-4. The Brewers scored their runs on a throwing error, a delayed double steal, a wind-blown popup that fell in shallow center field, a fielding error on that same play, a wind-aided triple and a suicide squeeze bunt, all in three innings of at-bats.

The triple belonged to Caleb Gindl, who motored to third after Rangers center fielder Craig Gentry crashed into the wall, forcing open a large gate. Gentry and left fielder Conor Jackson worked together to close it so play could continue.

"It was crazy out there, " Gindl said. "it was scary in the outfield. After a while we were all just playing deep, knowing the ball would either get to us or blow out. "

Play didn't last long after the Brewers' four-run third inning. Brewers reliever Manny Parra pitched a scoreless fourth, then the grounds crew covered the field before the bottom of the inning could begin.

After a delay of just 12 minutes, the game was called.

First of all, I absolutely LOVE the fact that Caleb Gindl uses two Old Norse words to describe the weather conditions during the game. It provides a certain primal, unhinged quality to the situation and adds a third element — nature — to the

contest. I also like the use of the onomatopoeic terms “pop” and “crash” because they serve to underscore the action.

The passage itself is a little lighter on the French and Latin roots than some of the earlier paragraphs and many of the terms are fairly short — the average word length comes in at about 4.6 characters. Some of this may be due to the fact that it is an online article (and attention spans are short) but it may also relate to the simple concepts at the core of the game itself. Words like “bat” and “ball” are very similar to their proto Indo-European roots (**bhat*- and **bhel*- respectively), suggesting that any associated activities are pretty basic to the language. Also, the sheer number and variety of numeric references (e.g. “three”, “third”, and “triple”) bring in many simple terms.

Updates:

- **5/6/2020** – <https://getpocket.com/explore/item/why-grown-ups-keep-talking-like-little-kids>
- **9/18/2019** – <https://getpocket.com/explore/item/english-is-not-normal>

114 thoughts on “Visualizing English Word Origins”



Douglas Harper says:

April 6, 2012 at 9:58 pm

That's a really intriguing use of the material. Congratulations.

[Reply](#)



santiago says:

May 4, 2012 at 11:34 am

Mike, great work, I've enjoyed it a lot. Some time ago I've made an analysis with words coined by Shakespeare that, being much more simpler than yours, I think it's related:

<http://blog.impure.com/?p=1050>

Also, I saw Douglas Harper was the first to comment your post, what if you propose him to create an api with its dictionary?

Thanks!

[Reply](#)



mkinde says:

April 19, 2012 at 11:19 am

I found this Winston Churchill quote online and thought it was appropriate: "Short words are best and the old words when short are best of all."

[Reply](#)



John says:

April 25, 2012 at 11:49 am

Just a quick correction:

I recognise the Dickens passage and it is NOT from 'A Tale of Two Cities'. It is, I believe, from 'Great Expectations'.

I ommit snarky comment about your eruditeness 😊

[Reply](#)



mkinde says:

April 25, 2012 at 1:28 pm

Ah, John, you are right! 'Tale of Two Cities' was where I started ... forgot when I switched texts. I will make the change. My apologies to all English majors everywhere (including you, mom).

[Reply](#)



David says:

April 28, 2012 at 7:56 am

Wouldn't it be more erudite to use "erudition" instead of "eruditeness"?

No snarky comment here.

[Reply](#)



Julie says:

May 2, 2012 at 8:51 am

Not to mention a snarky comment about the correct spelling of "omit"...

[Reply](#)



Peter says:

April 25, 2012 at 11:53 am

Very cool! I worked on a related project a while back and found it helpful to augment EtymOnline with data from AllWords. The project was to provide an Etymological Thesaurus so you could search for different ways of saying something and plot the results by age and language of origin. Doesn't look like any of the screen-scraping code is working anymore; not surprising as it was quite a while back...

[Reply](#)



Earl Killian says:

April 25, 2012 at 12:00 pm

I bet a lot of people would love to experiment with this, even if it is a somewhat manual process. How about creating a webpage where people can upload text, and you analyze it?

One idea: how about run your program on the examples in Orwell's Politics and the English Language essay?

[Reply](#)



mkinde says:

April 25, 2012 at 1:30 pm

Earl ... that was my original intent but I ran low on time. It is definitely on my 'to do' list. Also, I'll see what I can do about the Orwell passage. Keep checking back!

[Reply](#)



mossby says:

April 26, 2012 at 7:25 am

I would like to second this request. I would use such a tool all the time.

[Reply](#)



M Burden says:

May 1, 2012 at 9:30 am

I wrote a basic system that does automatic etymological highlighting, if you're want to check it out:

<http://huco.artsrn.ualberta.ca/~mburden/project/threads.php?thread=Shakspere>

It uses the same source for etymologies.

The parsing isn't very intelligent, so its liable to make mistakes, but it can give a flavor of what the result would be like.

Mike: It's really great to see someone else trying this.. and I'm glad you have the patience to do each text manually 😊

[Reply](#)



mkinde says:

May 1, 2012 at 8:48 pm

Your site seems to work very well. Certainly a great place to start for anyone interested in trying a few samples.

(BTW: it's either patience ... or stubbornness.)

[Reply](#)



Nathan says:

April 12, 2020 at 2:54 am

Hey! I am very interested in this etymology program if it is still up and running? I think the link has died??

[Reply](#)



mkinde says:

April 23, 2020 at 8:23 pm

It was all done using a very old MS Access program back in 2012. I have not revised the process using new tools ... might be about time.

[Reply](#)



SeriousSham says:

April 25, 2012 at 2:36 pm

This is supremely cool, and I suspect it might prove very, very valuable insights for some text analysis tasks – at least as a starting point.

I hope with all my heart that you will find some time to release the code and/or build a usable input page, because this is fun and useful in equally huge measures.

[Reply](#)



Brian Hook says:

April 25, 2012 at 3:34 pm

Very enjoyable. But why distinguish Anglo-French, Old French, and Middle French from Medieval and Classical Latin? Those words are largely based on Latin. (Why distinguish medieval from classical Latin, for that matter?) If you look at the Monty Python example, "supreme" and "aquatic" seem to be Anglo-French, but *supremus* and *aqua* are perfectly normal Latin words (full disclosure: I teach this stuff). The good news, for word-enthusiasts, is that these etymologies are even

deeper and more complicated than this reveals. But the fact that many words arrived into English having percolated through several languages (e.g. Greek-Latin-French) is obscured here.

[Reply](#)



mkinde says:

April 25, 2012 at 4:11 pm

Hey, Brian, you bring up a great point. I struggled with finding an easy way to present the deeper history of individual words and my tentative solution was to provide a hyperlink to Douglas Harper's dictionary for each word or word segment. Kind of a cop out, I know, since my intention was to show this entire process visually. I suppose it might be interesting to have a full etymology pop up when you mouseover the text. You might even be able to do a map! As far as distinguishing between the varius word origins listed, I must admit that this was based on what I could do programmatically. I basically searched for the first language mentioned in Mr. Harper's definition knowing that this might introduce some errors. My hope that this would at least provide a starting point for the non-expert. Thanks for your thoughts.

[Reply](#)



James says:

March 6, 2020 at 12:20 am

These words are Middle French, not Anglo-French. The proximate language of borrowing is indicated because that goes a long way to explaining the cultural reasons for the borrowing as well as the phonetic shifts the etymon had undergone by the time of borrowing. An overwhelming majority of the words here are from Proto-Indo-European—if you want to go back to an earlier originator, why not just indicate PIE and call it a day? Furthermore, choosing the proximate language explains sound shifts but conceals the “percolation,” whereas choosing the original language also conceals the later “percolation,” all while leaving us in the dark with respect to the sound changes and cultural impetus for the borrowing.

[Reply](#)



mkinde says:

March 6, 2020 at 7:41 pm

Thanks, James. There's definitely an opportunity to revisit this project with some of the tools that have been developed over the past few years. Let me give it some thought.

[Reply](#)

Pingback: [Visualizing English Word Origins | Dave Schumaker](#)



Carl says:

April 25, 2012 at 5:29 pm

This is cool, but there are some real limitations to using colors for this (beyond the obvious problems of varying qualities of reproduction on various screens). For the color-blind among us, distinguishing between some colors (and it's different depending on your color-blindness) is maddening or impossible, and it's very hard to "map" the colors — to remember that this shade of orange represents Middle French, for instance. Not meant to be a negative criticism, just something to keep in mind. I don't know if a mouseover could easily present the relevant derivations so that the color isn't the only source of information.

[Reply](#)



mkinde says:

April 25, 2012 at 8:30 pm

Agreed. Color is a tough variable to tackle and I'm not sure how successful I was. I did try and use the mouseover option but it may not work the same on all devices.

[Reply](#)



fred says:

April 25, 2012 at 8:25 pm

fascinating!

[Reply](#)

Pingback: [Off topic: Colorful word origins, WiFi for poop, beetles at the museum, musical cities | Good Morning Silicon Valley](#)



Ed. Floden says:

April 26, 2012 at 1:53 pm

It's all grey to me.

[Reply](#)



Henry Norr says:

April 26, 2012 at 2:08 pm

What's the book on the development of the English language that inspired this interesting project?

[Reply](#)



mkinde says:

April 26, 2012 at 10:05 pm

The book is called 'The English is Coming!' by Leslie Dunton-Downer. I would also recommend 'The Story of English' by Robert McCrum, et. al. as well as 'Words and Rules' by Steven Pinker.

[Reply](#)



TonyB says:

April 26, 2012 at 9:57 pm

Your initiative is fascinating and the web site a great idea. Perhaps you could charge for every analysis. If you need the money, keep it, if you don't, have it go straight to your (or the user's) favorite charity.

[Reply](#)



sedeer says:

April 28, 2012 at 9:19 am

That's a really great idea! There are a couple of things I'd like to try out and maybe do differently. Would you be willing to share a copy of your script so I don't have to start from scratch?

[Reply](#)



Doris says:

April 28, 2012 at 10:21 am

You're right about the "uninitiated" not understanding the baseball language. You could, for that matter, be talking cricket.

[Reply](#)

Pingback: [Miscellany 31 - Hieropraxis | Hieropraxis](#)



Elyse Hendricks says:

April 30, 2012 at 11:49 am

This would make a great teaching tool for LIN 101 students or high school seniors. If you could make it interactive, they could try to color-code and it could be a game or exercise. It could be fun (and illuminating) to have them start with broad origins—Latin/OE/Greek—and then show the distribution channel (so to speak).

[Reply](#)



Sonia Lal says:

April 30, 2012 at 6:37 pm

Really interesting post. There are more old English origin words than I thought there could be.

[Reply](#)



Laurel says:

April 30, 2012 at 6:48 pm

Great stuff! Would like, I mean love, to have an app that did this. I do not agree with dropping Medieval Frecnh, etc., however. Part of the fascination of words is where they came from and how they evolved to have their current meaning.

[Reply](#)



Ian says:

April 30, 2012 at 10:06 pm

Pity about Latin and Greek being the same color (to my eyes). Lime green isn't taken, why not assign it to Greek?

[Reply](#)



mkinde says:

April 30, 2012 at 10:44 pm

Good idea and easily done.

[Reply](#)



Joseph says:

May 1, 2012 at 3:21 am

You should run this analysis on the text of your article 😊

[Reply](#)



Chris Hunt says:

May 1, 2012 at 3:52 am

What a great idea!

I would query your identification of "Joe" as american English though. When used to mean "coffee", it's certainly American, but as a shortened form of Joseph it seems pretty English – e.g. the clown Joe Grimaldi b.1779

If you follow the precedent of your treatment of "Tom", though, you should mark "Joe" as being Hebrew.

[Reply](#)



mkinde says:

May 1, 2012 at 12:18 pm

Good catch ... and you expose one of the big barriers to creating a fully-automated app. Interpreting these designations is a mostly manual affair at this point.

[Reply](#)



mkinde says:

May 1, 2012 at 12:39 pm

Corrected.

[Reply](#)



Eric says:

May 1, 2012 at 10:10 am

This is fascinating, but isn't plural 's' morphology rather than etymology? If you are going to pick out suffixes then you should be splitting the words up and looking at the derivation of their morphemes: -able, -less, -ly, -hood, -ing, etc.

[Reply](#)



mkinde says:

May 1, 2012 at 12:29 pm

That was a tough call. If the word was listed as a stand-alone entry in Douglas Harper's site, I didn't split it up into sub-components. However, if the word was pluralized or involved some sort of verb ending, I tried to make the distinction. There were certainly some interesting dinner conversations during construction.

[Reply](#)



Warren Rempel says:

May 1, 2012 at 11:11 am

I agree with the other commenters who feel this is worthy of sharing. I also agree that you deserve some financial compensation for your efforts.

This calls for an "app"!!

[Reply](#)



mkinde says:

May 1, 2012 at 12:43 pm

Unfortunately, it is more a "proof of concept" than a functioning program at the moment. Your interest is noted!

[Reply](#)

Pingback: [Word-by-Word Pictures of the History of English |](#)



Tim Breen says:

May 1, 2012 at 1:59 pm

In the Dickens text I noted when hovering over words that some are tagged as "American English" (marbles, door-knob) and one is tagged as "United States" (fire-cracker). What's the distinction?

[Reply](#)



mkinde says:

May 1, 2012 at 8:43 pm

No distinction ... my error.

[Reply](#)

Pingback: [NewLink: Ideas Illustrated » Blog Archive » Visualizing English Word Origins | Leonieke Net](#)



mkinde says:

May 1, 2012 at 8:39 pm

Anyone itching to play with their own text samples should try out this site from Michael Burden:

[http://huco.artsrn.ualberta.ca/~mburden/project/threads.php?
thread=Shakspere](http://huco.artsrn.ualberta.ca/~mburden/project/threads.php?thread=Shakspere)

He cautions that it is only about 80% accurate but it is still worth a look.

[Reply](#)



Jestyn says:

May 2, 2012 at 3:31 pm

Hope you saw that the Economist magazine's Prospero blog wrote about and linked to you?

[http://www.economist.com/blogs/johnson/2012/05/etymology?
fsrc=gn_ep](http://www.economist.com/blogs/johnson/2012/05/etymology?fsrc=gn_ep)

[Reply](#)



Jestyn says:

May 2, 2012 at 3:33 pm

oops! That'll be Johnson blog, Prospero is the arts one, Johnson is the words one...

[Reply](#)



mkinde says:

May 2, 2012 at 8:04 pm

Thanks for the heads up! Missed it.

[Reply](#)



Brody says:

May 2, 2012 at 7:51 pm

Not read all the comments, but for me, the most interesting use is to use track whether and how the extent of etymological diversity shows a correlation to educational level and social background. Do the 'New York Times' and 'The Guardian' show larger levels of Latinate (French/Latin) to Germanic (Saxon/Norse) than say 'The Sun' and 'National Enquirer'.

And if so, does that indicate that Norman/Anglo-saxon social strata distinctions since the Norman conquest haven't blurred to the degree we might think.

[Reply](#)

Pingback: [More Coffee Please » Neat Stuff from Elsewhere Thu May 03, 2012](#)



Andrew says:

May 3, 2012 at 8:49 am

Would be great to be able to cut and paste text and scan.

[Reply](#)



Colleen says:

May 3, 2012 at 9:18 am

The differences in the shades of orange aren't really noticeable on my monitor. How about purple?

[Reply](#)

Pingback: [Ideas Illustrated » Blog Archive » Visualizing English Word Origins » Mile Wide... Inch Deep...](#)



Andy RYLAND says:

May 4, 2012 at 3:02 am

Very interesting, I wonder though if two things are going on here. In teaching dyslexics it has been found that different colored overlays or paper can make the reading easier so for a proper test of readability we may need to swap colors around.

However perhaps there is also a connection here with writing in Plain English which seems to show some connection with old English. see <http://www.plainenglish.co.uk/> What do others think.

[Reply](#)



Harold says:

May 4, 2012 at 9:24 am

There was a “language purism” movement in the nineteenth century touting “Anglo-Saxon” words as somehow more masculine and almost racially superior to decadent and effeminate French and Latin.

Actually, this probably goes back to polemics about translations of the Douay–Rheims Bible (Catholic, 1582), which was deliberately, reactively Latinate. In short, there are both positive and negative aspects to all this. I for one, regret the obliteration of Latin. A book by Kenneth Haynes, “English Literature and Ancient Languages (Oxford, 2003) reminds us that quite a bit of “English” literature was written in Latin — Milton and so on — and has now become quite lost to us.

[Reply](#)

Pingback: [Visualization of the Week: The origins of English | Share Blog](#)



Rory Van Tuyl says:

May 5, 2012 at 12:58 pm

I wrote a piece about this and obtained pretty much the same results: 75% OE+ON; 25% other.

This despite the fact that the dictionary has it the other way 'round: only a quarter of words listed come from Germanic sources.

Also, the Latinate composition of American news writing has almost doubled from the 18th century to the present.

See:

<http://www.roryvantuyl.com/PDFs/My%20Excellent%20Etymological%20Adventure.pdf>

[Reply](#)



mkinde says:

May 5, 2012 at 6:00 pm

Thanks for the link. That's a great paper ... worth the read for anyone looking for more word sample analysis. The trend in newspaper writing is interesting. At first I was thinking the drop in OE & ON words since the 1900s would be due to new words associated with new technology (e.g. computer, internal combustion). That doesn't seem to be the case, though.

[Reply](#)



AnWulf says:

August 10, 2012 at 7:03 pm

Fetching writ but it has some mistakes. A few small ones like potato highlighted as a Romance Language word. (It came thru Spanish but is Carib which got it from farther south .., papa ... which is the nowadays Spanish word for potato as well.) And ocean came thru French from is Greek.

The bigger mistake was ... "But Old English was the language of a relatively primitive people, and therefore lacked root words from which certain sophisticated concepts could be expressed."

This is utterly wrong.

You giv byspels of: council [12th c]; nobility [13th c]; city [13th c]; power [14th C]; community [14th C]; civil [14th c]; government [1550]; republic [1600].

council – OE was witan, moot, gemoot, þinge, þingþ (burhgeþingþ was a city council) and more.

nobility – OE was æþelu, æþelnes (athel)

city – OE (and today's English) burg

power – mægen

community – sundry words hinging on the meaning: gemænnes, gemænscip, gemāna, asf ... nowadays neighborship also stands

civil – burg/burh ... burglagu = civil law

government – lēodweard, gerec, wealdung, (ge)wissung

republic – folcagende (ruling [folk ruling]) or folcscip (folkship)

Other words that you show could eathly hav been made from OE roots ... BTW, æschere and scipcræft both mean naval force, flotlic meant nautical/naval.

Most of the words crafted from Latin or Greek roots were done by choice, not by needs.

[Reply](#)



Luc Gauvreau says:

May 5, 2012 at 8:06 pm

Bonjour!,

Fascinant, félicitations pour cette superbe visualisation! It's not "déjà vu", it's "jamais (never) vu"!

Great idea for any languages, langues, lingua...

Merci beaucoup (from Montréal, Canada)

[Reply](#)



David says:

May 6, 2012 at 11:33 am

Medical Greek missed:

anatomic- (and -ic- elsewhere) | arthritis | centr-(e) | -graph |
mechanic- | symptomatic |; even | axis | has Gk | axon | behind it.
Not a game for the casual player?

[Reply](#)



John M says:

May 7, 2012 at 1:47 am

I agree with David. Too many Greek words were marked as Latin. Douglas Harper's dictionary is accurate (e.g. "from L. mechanicus from Gk. mechanikos") but the user of the dictionary wasn't. The Just the fact that the Romans borrowed Greek words doesn't make them "of Latin origin". An amateur attempt.

[Reply](#)



Yannis says:

May 8, 2012 at 8:33 am

I agree. The problem is that an etymological dictionary is intended to be used by people with some understanding of etymology, not by some automatic process. Greek words that got transliterated to Latin in western European translations of Greek texts will have this attestation recorded in their etymology for historical reasons. A philologist will understand this, while an automatic system will erroneously mark them as Latin in origin.

[Reply](#)



mkinde says:

May 8, 2012 at 12:51 pm

So a distinction is made between different language transmission methods, with an emphasis on more "natural," peer-to-peer adoption of words? In other words, Latin is different from Latin Translation? I can see that. Thank you for the constructive criticism.

[Reply](#)



mkinde says:

May 8, 2012 at 12:39 pm

Of course I'm an amateur ... that was the whole point of the exercise.

[Reply](#)



mkinde says:

May 7, 2012 at 12:06 pm

For good or ill, I used a “last in, first out” policy which favored recency. If Douglas Harper’s dictionary said that the word was of Greek origin but came to English through a Latin derivative, I credited Latin.

[Reply](#)



Lou says:

May 7, 2012 at 6:04 am

“Your browser does not support charts”

Not sure what I’m missing on the iPad.

[Reply](#)



mkinde says:

May 7, 2012 at 12:17 pm

Possibly a problem on my end with the Google Charts API. Works on my wife’s iPad, though, so it may be some conflict. Thanks for the heads up.

[Reply](#)

Pingback: [Found while foraging \(May 7, 2012\) « Inspiring Science](#)



Sean says:

May 8, 2012 at 3:55 am

This is great! I tried to visualise the online etymology dictionary by looking at paths individual words had taken through different languages. I found that the origins of words are much more graph-like than tree-like. There’s an interactive graph here:

<http://theadventuresofauck.blogspot.de/2009/11/etymologies.html>

[Reply](#)



Yannis says:

May 8, 2012 at 8:54 am

The reason there is not much tagged as Middle English (ME) in origin is that ME is a development stage between Old English (OE) and Modern English (MnE). The ME vocabulary will to a great extent consist of words from OE and words imported from Norman French; both of these would be cited with those respective origins unless they underwent significant morphological change in the ME period, or are unattested elsewhere.

[Reply](#)

Pingback: [blog_riley_|_DETERIORATA XII:2](#)

Pingback: [5 Interesting Links for 5-11-2012 | Tales to Tide You Over](#)



Nancy Lebovitz says:

May 12, 2012 at 6:31 am

Thanks for doing this.

It might be interesting to apply your approach to the dialogue from The Council of Elrond– in _The Road to Middle-Earth_ (p117-122), Shippey shows how Tolkien varies the speakers' vocabulary and way of speaking to show their culture and age.

[Reply](#)



mkinde says:

May 12, 2012 at 6:02 pm

That would be an interesting challenge! I would have add categories for the various Tolkien languages. He documented everything so well that there must be an etymological dictionary of Elvish out there somewhere.

[Reply](#)



N Baker says:

May 12, 2012 at 7:47 am

Great work and shows that English is still mostly English despite a thousand years of additions.

[Reply](#)

Pingback: [Listen & Learn](#) » [Blog Archive](#) » [Colour coding languages](#)



Pepijn de Vos says:

May 13, 2012 at 7:31 am

How did you determine the origin of a word? I tried to do some Dutch texts.

If I look at the etymologic dictionary, a lot of words have origins in several varieties of dutch, german, french, english, latin, greek and more at the same time.

[Reply](#)



mkinde says:

May 14, 2012 at 7:48 pm

You definitely have to set some ground rules when you choose an origin to emphasize ... and not everyone will agree with your choice. I based my display on the first source mentioned in the etymology dictionary (usually the most recent) and then made adjustments from there. It's tricky since some words have multiple meanings with different origins.

[Reply](#)



adastrax says:

May 15, 2012 at 3:28 pm

The extend of OE words in use is impressive, considering that 95% of the OE lexicon was lost after the Norman Conquest. What remains of OE in ModE is mainly the root of the language, grammatical words and low-register lexemes.

What I wonder though is what your definition of Anglo-French, Old French and Middle French is. As far as I know there were only two kinds of French which had an influence on the development of the language: Norman French (with the Normans) and Central French (with the Plantagenets).

Maybe this would be easier to understand if you assigned time periods to the languages.

[Reply](#)



mkinde says:

May 15, 2012 at 8:34 pm

Douglas Harper covers some of the specifics on his site (<http://www.etymonline.com/abbr>). He breaks French down into several time periods, starting with Gallo-Roman (c. 500-900 C.E.) which is an intermediary between Vulgar Latin and Old French. Old French (c. 900-1400) is followed by Middle French (c. 1400-1600), and then French (presumably in its modern form). Old North French and Norman are called out separately, as is Anglo-French — the French used in England from the time of the Norman Conquest (1066) through the Middle Ages. I find this last one an interesting distinction.

[Reply](#)



AnWulf says:

August 10, 2012 at 7:18 pm

@adastra ... “considering that 95% of the OE lexicon was lost after the Norman Conquest.”

I don’t think this is true at all. Even if one tallies up the words that hav fallen out of noting since after Middle English, most of OE is still found. We’ve stopp’d noting a lot of the words, but they are still there.

[Reply](#)

Pingback: [Visualizing English Word Origins | Nobil Group](#)



Jules says:

May 19, 2012 at 3:51 pm

A mistake seems to have cropped in to your analysis — you have ‘marbles’ down as an American English word, but if you read the description it links to, it only suggests that the sense of “mental faculties” is of American origin, which isn’t the sense it’s being used in here... instead you need to follow the link “plural of marble” and use the origin from there (old French).

[Reply](#)



mkinde says:

May 20, 2012 at 9:51 pm

Good catch. I’ve made the change.

[Reply](#)

Pingback: [Etymologies at a Glance | Samuel Arbesman](#)



Russell Cross says:

May 26, 2012 at 9:01 am

A splendid and entertaining way to represent etymologies. Sure, it’s not perfect, but as you say, this encourages people to go “digging in to the full etymology site to explore the full history of each word” then your work is done 😊

[Reply](#)



mkinde says:

May 27, 2012 at 1:34 pm

Does that work as a CYA?

[Reply](#)



Andy T says:

May 26, 2012 at 10:20 am

Do you have a study on words from language systems outside of the European context. African, Native American or Asian examples? Your work is phenomenal.

[Reply](#)



Skip Barnett says:

June 10, 2012 at 10:23 pm

Really fascinating! Who would've guessed that Old English rooted words would still be such major players in all these texts, even in the legal and medical texts! I would've assumed a much stronger showing for Latin, Greek and French words than you found and Old Norse, too. And thanks to a former student – Mr. John Munley – for pointing this out to me! I'll torture future students with it.

[Reply](#)



Leslie Dunton-Downer says:

June 13, 2012 at 11:36 pm

As a big fan of visual tools, I really enjoyed looking at your website. And I agree with others who see an exceptional 'word origins & ratios of borrowing' app in the making for analysis of texts in English. Why not approach editors of the Oxford English Dictionary and propose to work with them to mine in real time their unique and rapidly evolving database? One thought in case you take this another step: it could be illuminating to present a variation that would offer the 'whole picture' pie chart take on each text alongside a pie chart based on the same text minus the top ten or so most frequently used words in Modern English. Or the second chart could devote one slice of the pie to 'x

number of most-used words' so you could see how these stack up overall, and how different the origin ratios can be with these words removed from the picture. As English-speakers continue to drop use of some of these words, this could also become a way to track an important trend in the language's evolution. Even with the examples you explored in your post, it could be worthwhile to see, for instance, which native English words are left in the Medicine text vs. the Sports text once most-used words are removed or put in a distinct category. To determine which words would make the 'most-used' cut, you could look at frequency of use data for Modern English and see where a significant drop occurs. I recall some slight variations even in top-ten most-used word lists based on different kinds analyses, so one approach could be to take only those words that appear at the top of all authoritative lists — maybe there are 8 or 9. For now, they are all of Old English origin. To put the role of most-used words in perspective, I note this point made on the "Most common words in English" page of Wikipedia: "The Reading Teachers Book of Lists claims that the first 25 words make up about one-third of all printed material in English, and that the first 100 make up about one-half of all written material." One upshot is that even small variations in use of high frequency words would significantly change borrowed word ratios.

Congratulations on your captivating visual work!

[Reply](#)



mkinde says:

June 14, 2012 at 2:03 pm

Thank you, Leslie! I enjoyed your book and hope that I was able to steer a few more readers your way. I'll definitely take your advice and contact the folks at the OED to see if there are some opportunities to work together. Your thoughts on most-used words also make a lot of sense. Establishing a "core" category would help shed light on the use of less common English words while emphasizing the idea that most of our every-day conversations involve a fairly narrow range of words.

[Reply](#)

Pingback: [What colour is your document – etymology visualisation](#) « [ffeathers](#)

Pingback: [Visualization of the Week: The origins of English - O'Reilly Radar](#)



Garrigus Carraig says:

August 3, 2012 at 10:28 pm

Very cool. 'Tom' ('Thomas') is not from Greek, rather from an Aramaic word meaning 'twin', but I suppose that's Etymonline's problem.

[Reply](#)



mkinde says:

August 6, 2012 at 1:58 pm

It is amazing how many refinements people bring up. I think we need a wikipedia for etymology. Thanks for the post!

[Reply](#)



Ben says:

August 15, 2012 at 10:55 am

The same correction offered about the name "Tom" could be said for "Ben" – a shortening of the old Hebrew name Benjamin, or Ben-yamin, meaning "son of the right hand" or "son of blessing." The prefix "Ben" is a common Hebrew one, and actually has its cognate in modern Semitic languages (i.e. "Osama Bin-Laden" where "Bin" signifies "son of"...). I believe your classification of the name "Ben" as Gaelic in origin is perhaps, then, a mistake – unless, of course, the character in question is Gaelic or Irish and is only named "Ben" (not a nickname).

I remember learning about the meaning of my name in an Ancient Hebrew course in college and really enjoying its literal meaning. To say that one "sat at the right hand" of someone was to say they had been given an inheritance or a blessing. Thus, Ben-yamin, "Son of Right Hand."

As a language nerd, this project astounds and delights me. Thank you for your work.

[Reply](#)



mkinde says:

August 21, 2012 at 9:08 am

The history of names seems particularly convoluted ... might merit a separate post. Thanks for your thoughts!

[Reply](#)

Pingback: [linguistische wochenendlektüre « lexikographieblog](#)

Pingback: [Ideas Illustrated » Visualizing English Word Origins | Data is Beautiful | Scoop.it](#)



zé do rock says:

August 11, 2012 at 3:31 am

jou kan also form long textes dat ar perfekt comprensible voor de englisch en do niet heff a single englisch woord in het, as diese sentence. of course jou kan do het mit meer aise wenn jou use also germanique dialectes as low german en afrikaans.

[Reply](#)

Pingback: [Ideas Illustrated » Visualizing English Word Origins | 6-Traits Resources | Scoop.it](#)

Pingback: [Ideas Illustrated » Visualizing English Word Origins | abcdef | Scoop.it](#)

Pingback: [Visualizing English Word Origins blog \(Kathryn\) « jkrenglish](#)

Pingback: [Ideas Illustrated » Blog Archive » Visualizing English Word Origins | IELTS, ESP and CALL | Scoop.it](#)

Pingback: [Visualization of the Week: The origins of English - Strata](#)



dw says:

March 11, 2013 at 2:16 pm

“Baseball” should not be classified as “American English”. To quote Wikipedia:

The earliest known reference to baseball is in a 1744 British publication, A Little Pretty Pocket-Book, by John Newbery. It contains a rhymed description of "base-ball" and a woodcut that shows a field set-up somewhat similar to the modern game—though in a triangular rather than diamond configuration, and with posts instead of ground-level bases.

[Reply](#)

Pingback: [News Roundup: Food, Culture, Taxes, & Telecommuting | Wherever I am, you are there also](#)

Pingback: [#writing Visualizing English Word Origins. The original source is ideasillustrated.... and it has some really interesting breakdowns of different types of writ... - English Cheat Sheets](#)

Leave a Reply

Your email address will not be published. Required fields are marked *

COMMENT *

NAME *

EMAIL *

WEBSITE

[Post Comment](#)

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

Search

Categories

[Design \(15\)](#)

[Education \(1\)](#)

[Favorites \(5\)](#)

[Infographics \(14\)](#)

[Information \(44\)](#)

[Programming \(3\)](#)

[Uncategorized \(1\)](#)

[Urban Planning \(1\)](#)

META

[Log in](#)

[Entries feed](#)

[Comments feed](#)

TAGS

American Film Architecture

Bias **Charts** D3

Data Database

Data

Visualization

Design Principles

Disinformation Dog

Donald Trump Egypt

Filter Graphs Industrial

Design Infographics

Information

Design

Information

Overload John Boyd

Journalism Knowledge

Management List Map

Media Obama Occupations

OODA Loop Organization

Politics Privacy Propaganda

Signage Simplicity Snow Day

Social Media Symbols

Tableau

TV Remote

Unemployment Usability

User Experience

Visualization Words

Popular Posts

Visualizing English Word

Origins

62.9k views

Most Popular Word Roots
in U.S. Place Names

28.3k views

Trends in NFL Football
Scores (Part 1)

17.8k views

How to Build the Perfect
Data Science Team

15.7k views

Propaganda vs.
Disinformation vs. News

15.4k views

Channel Surfing Ain't
What it Used to Be

15.2k views

A Force Node Diagram of
the U.S. Interstate
System

13.6k views

Three Rules of PowerPoint

7.3k views

Thick as a [LEGO] Brick

6.1k views