

# Writing good English: is scientific English a Latin language in disguise?

Mauricio Rocha-e-Silva<sup>1</sup>

<sup>1</sup> Universidade de São Paulo, São Paulo, Brazil

**BACKGROUND:** English is the lingua franca of science; it is the language of the two last world superpowers and the language of four out of the world's ten greatest producers of science; it is a fairly simple language and the most hybridized language in history, with Latin and French contributing 60% of the entire English lexicon. The object of this study is to determine whether the frequency of use of imported words is a function of literary genre.

**METHOD:** Texts were randomly selected from (a) medical scientific original articles, (b) newspaper financial reports, (c) sport reportages, (d) literary texts and (e) colloquial English; for comparison a collection of similarly distributed texts were selected from Portuguese; the frequency of occurrence of Latin or Neo-Latin words was determined in the English texts as well as the occurrence of non-Latin or non-Neo-Latin words in the Portuguese texts; a one-way analysis of variance was used to determine whether significant differences occurred between genres in the two languages.

**RESULTS:** The frequency of occurrence of Latin/French words in English text was significantly dependent on the literary genre, being maximal in medical scientific texts and minimal in colloquial English; in contrast, the frequency of occurrence of non-Latin words in Portuguese was constant throughout the same literary genres.

**CONCLUSION:** The use of Latin/French words in English is directly proportional to the complexity of the literary genre, a phenomenon not observed in Portuguese, a typical Neo-Latin language.

**KEYWORDS:** Medical Education, Scientific language, Ethymology

Rocha-e-Silva M. Writing good English: is scientific English a Latin language in disguise? MedicalExpress (São Paulo online). 2018;5:mf18001

Received for Publication on November 20, 2017; First review on December 18, 2017; Accepted for publication on January 3, 2018; Online on January 19, 2018

E-mail: marosilva36@gmail.com

## INTRODUCTION

Over the last century English gradually became the lingua franca of science, more specially of the so called hard sciences. This has not always been the case. A century ago, German, French and Italian, amongst others, were totally valid languages for transmitting scientific novelty. After World War II, things changed. Some very good reasons helped to bring this about.

1. English had been for over 250 years the language of two successive alpha world powers. The United Kingdom emerged with the industrial revolution in the middle of the 18<sup>th</sup> century, while the United States emerged as an industrial giant at the beginning of the 20<sup>th</sup> century. World powers export a lot more than merchandise, and language stands very high on their export agenda.

2. English is also the native language of four among the top ten contemporary producers of science,

namely the USA, the UK, Canada and Australia. Together, they account for 30% of all published original scientific articles.<sup>1</sup>

3. English is a relatively simple language in terms of syntax. Nouns are very predominantly neuter; the very few exceptions are given names, some titles, such as king/queen and a very few animals, such as horse/mare in the domestic, tiger/tigress in the feral domains. English articles and adjectives are universally neutral. Regular verbs belong to a single conjugation which contains only four variant forms; 370 irregular verbs follow a fairly limited number of patterns. Phrase structure can be kept very simple: the most common phrase sequence is subject verb object(s).

4. English is probably the most intensely hybridized language in the history of human communication. It is probably the only major living language where imported words are more numerous than the words belonging to the original family trunk. Figure 1 shows the distribution of approximately 80,000 words by origin

DOI: 10.5935/MedicalExpress.2018.mf.001



Copyright © 2016 MEDICALEXPRESS. This is an open access article distributed under the terms of the creative commons attribution Non-Commercial License ([creativecommons.org/licenses/by-nc/4.0/](https://creativecommons.org/licenses/by-nc/4.0/)) which permits unrestricted non commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

determined through a computerized survey of the Shorter Oxford Dictionary. According to this survey, approximately 60% of all words in the English vocabulary are of Latin origin, either directly or through French.<sup>2</sup> A different survey arrived at a similar distribution.<sup>3</sup> A very large number of Latin and French words have moved untranslated into English: vice versa, in flagrante delicto, cul-de-sac, a-propos are easily remembered examples.

Neo-Latin languages exhibit a completely different pattern: figure 2 shows the origin of the vocabulary of French and Portuguese, two of the five major Neo-Latin languages: Latin contributes most of the lexicon. Spanish and Italian have not been researched for this study, but it is self-evident that their vocabularies have equally predominant Latin origins. The 5<sup>th</sup> major Neo-Latin language, Romanian, encircled by Slavic speaking nations contains a relatively small proportion of Slavic vocabules.<sup>4</sup>

Apart from English, other non-Neo-Latin European languages contain sizeable contingents of words derived from Latin. The analysis of the lexicon of these is beyond the scope of this study. However, it can be safely stated that this large contingent of “borrowed” Latin words in all European languages is a direct consequence of the Renaissance, and of the Industrial and scientific revolutions. Almost every invention, discovery or concept introduced over the past 600 years received names constructed from Latin or Greek roots, with a predominance of the former.

But English stands apart from other European languages because of the French contribution. The Norman conquest of England in 1066 meant that for at least 350 years, French became the court and official language of the kingdom. It gradually evolved into Anglo-Norman, now a dead French dialect. But lines in the Canterbury Tales (written in 1380-1400)<sup>5</sup> are very telling:

There was also a nun, a prioress (...)  
 And she was called Madame Eglantine (...)  
 She spoke French well and very stylishly  
 After the school of Stratford-at-the Bow  
 ‘Cos French of Paris was unknown to her<sup>5</sup>

This obviously contains a hint of sarcasm, Chaucer making subtle fun about Mme. Eglantine’s effort to show off; but it does tell us something about life in London by the end of the 14<sup>th</sup> century. Anglo-Norman was still alive and clearly differed from what Parisians spoke at the time. The next century saw the gradual death of Anglo-Norman, abandoned in favor of English. But its very existence left behind a long trail which is still with us. From William the Conqueror to the end of the 14<sup>th</sup> century, Old English gradually evolved into Middle English: the first Latin “invasion” came with a very strong French flavor. The second Latin invasion began soon afterwards and helped to turn Middle English into Modern English, the language now in use.

All of this is common knowledge, though some native English speakers, including educated ones, have a pretty vague idea about the size of the Latin “invasion”. The author has frequently come across persons who “had no idea that Latin was so pervasive in English”. We shall see that this lack of perception about the size of the invasion has roots in another probably unique feature of Modern English, which is the object of this communication: we shall examine the differential frequency of Latin and French imports into English as a function of literary genre.

## METHOD

Samples of text were collected from the following literary genres: medical scientific original articles, financial newspaper reports, sport reportages, literary texts, colloquial English, according to the following procedures: samples containing 250 – 600 words were randomly selected as described.

a) Medical scientific original articles: 20 samples were collected from scientific journals. Randomly selected articles (published between 2014 and 2017) were collected from Google Scholar in the following general medical categories: cardiology, dermatology, gynecology, nephrology, neurology, pediatrics, pneumology, obstetrics, oncology and orthopedics. Abstracts or fragments of the Discussion section were chosen from 2 articles from each of the above medical chapters.

b) Financial reports: 20 samples were collected from the following periodicals: The New York Times, The Washington Post, The Guardian, CNN transcripts; BBC transcripts. Four articles published in December 2017 were randomly selected from each these sources.

c) Sport reportages: 20 samples were collected from the same periodicals, published over the same time period

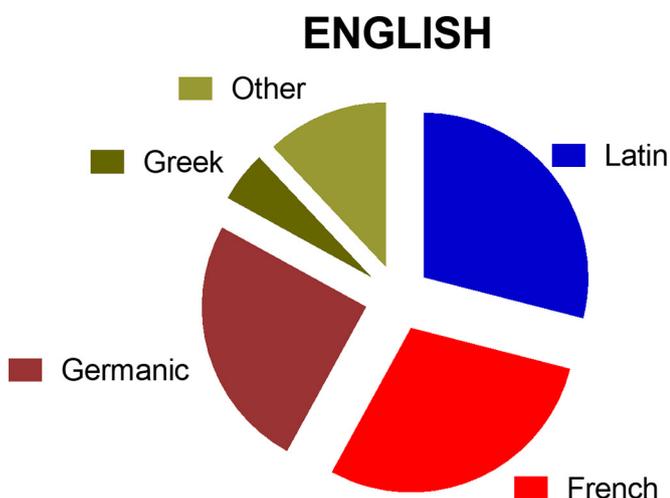
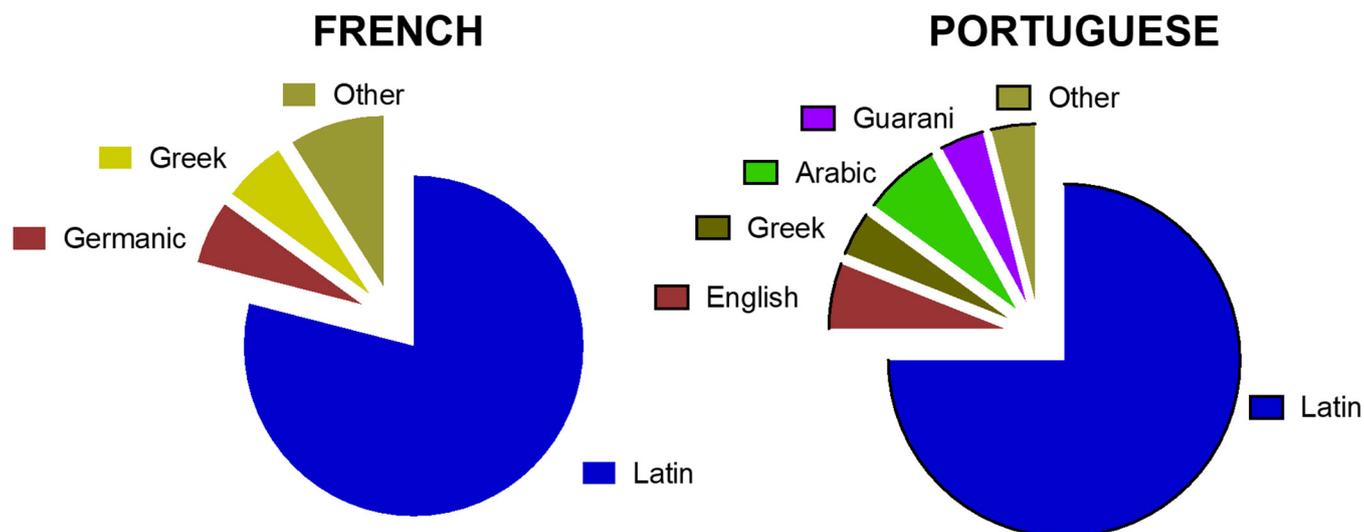
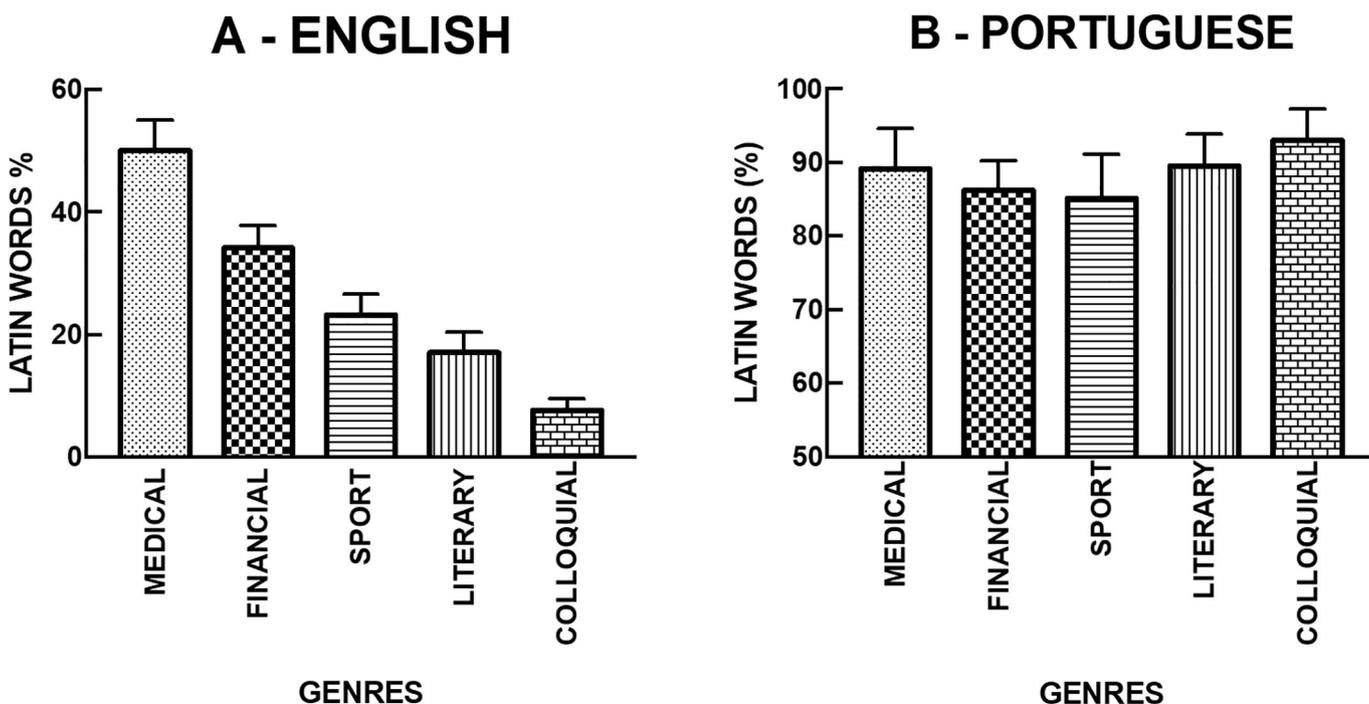


Figure 1. The percent distribution of words in Modern English as evaluated by a computerized survey of 80,000 words in the Shorter Oxford Dictionary.<sup>2</sup>



**Figure 2.** The percent distribution of words in Modern French and Portuguese as evaluated by estimating the frequency of occurrence of “borrowed” non-Latin or Neo-Latin words in each language.



**Figure 3.** Percent occurrence of Latin words in English and in Portuguese. The analysis of variance showed that both distributions conform to normality. The distribution in English is significantly related to genre ( $F_{4,95} = 419.5$ ;  $p < 0.001$ ), while the distribution in Portuguese showed no significant effect of genre ( $F_{4,25} = 2.4$ ;  $p = 0.075$ ).

relating to baseball (n=3), basketball (n=3), boxing (n=3), cricket (n=2), golf (n=2), hockey (n=2), soccer (n=3) and swimming (n=2). Baseball and hockey articles were selected exclusively from the American sources, cricket from the British sources.

d) Literary sources: 20 samples were collected from the following authors: Jane Austen (*Pride and Prejudice*), Hermann Melville (*Moby Dick*), George Bernard Shaw (*St*

Joan), Oscar Wilde (*The Importance of Being Earnest*), Mark Twain (*The Adventures of Tom Sawyer*). Four samples were collected from each publication.

e) Colloquial English: 20 samples were collected from the scripts of the following films: *Clockwork Orange* (Stanley Kubrick 1971), *From here to eternity* (Fred Zimmermann 1953), *Mighty Aphrodite* (Woody Allen, 1995), *Pillow Talk* (Michael Gordon, 1959), *Some like it hot* (Billy

Wilder, 1959), *The Apartment* (Billy Wilder, 1960), *When Harry met Sally* (Rob Reiner 1989). All stage directions were suppressed leaving only dialogues. Technical dialogues, especially legal arguments and technical descriptions were likewise omitted.

A similar, albeit limited procedure, was adopted for the same genres in Portuguese: 6 samples from each genre were collected: Portuguese abstracts of six scientific articles were collected from the SciELO collection;<sup>6</sup> financial and sports reportages were collected from two leading Brazilian news sources: *O Estado de São Paulo* and *www.globo.com*; literary samples were collected from writings of Machado de Assis, *Eça de Queiroz*, Fernando Pessoa, Vinicius de Moraes, Guimarães Rosa and Raquel de Queiroz. Colloquial Portuguese texts were collected from Facebook, but authors will remain anonymous for obvious reasons.

Each English sample was marked for words of Latin or Greek origin. Each Portuguese sample was marked for non-Latin language origins. The percentage of marked words within each text was established and means and standard error for each genre were determined.

Statistical analysis. A goodness of fit was performed and all five genre collections conformed with normal distribution. A one-way Analysis of Variance test was performed to compare the incidence of Latin/Greek words in each English genre. A similar procedure was performed to compare the incidence of non-Latin Language words in each Portuguese genre. Significance was assumed at  $p < 0.05$ .

## ■ RESULTS

Figure 3 shows the results for the incidence of words of Latin/Greek origin throughout the five genres in English and Portuguese. The very tight standard deviations for each genre in both languages clearly show that all of the selected genres use a highly coherent lexical architecture which is pervasive inside each genre. Panel A exhibits results for English. Differences between genres proved to be highly significant by one-way Analysis of Variance ( $p < 0.001$ ). Figure 3 also exhibits the equivalent results for Portuguese. In sharp contrast with what happens in English, all literary genres in Portuguese exhibit a very high proportion of words of Latin origin, with no significant differences between genres ( $p = 0.075$ ).

## ■ DISCUSSION

The essential finding of this study is shown in Figure 3: the frequency of occurrence of “borrowed” Latin/Greek words in English is a function of literary genre; in contrast, the use of “borrowed” non-Latin words in Portuguese is

independent of genre. The explanation for this is probably related to precision. The history of the French/Latin “invasions” shows that whenever a new level of linguistic precision became necessary, this precision generally required the use of “borrowed” Latin words. Neo-Latin languages, here exemplified by Portuguese, required precisely the same Latin words to express precision. Therefore, no word borrowing was required. Data for Figure 3, panel B came from my own native language because it would be easier for me to derive the required samples. But I can safely hypothesize that the same pattern would occur for any Nao-Latin language. In all of them, precision is mostly brought in by Latin derived words.

Even though imported words represent approximately 75% of the English lexicon, English speakers never use 75% of imported words in their spoken or written utterances. The reason for this begins to answer the question posed by the title. Normal texts contain at least 30% of crucial sentence-forming connective words (prepositions, conjunctions, articles, auxiliary verbs), which are never imported. In fact, it is virtually impossible to write naturally in any language without these vocabulary elements.<sup>7</sup> You might of course imagine examples such as, for instance “diligent people prefer coherent, functional solutions”, which is an all-Latin phrase, but any writer worth his salt would probably go for “most diligent people might prefer coherent and functional solutions”, where connectives represent 30% of the words.

Another point, this one raised at the end of the introduction, relates to the fact that native English speakers are usually surprised by the size of the Latin “invasion”. I believe this study offers an explanation: in most situations, people use colloquial or literary genres, precisely the genres where the “invasion” is minimal.

It would be interesting to look at the relative importance of the occurrence of “French” vs. “Latin” vocabules in this functional relation between lexicon and genre in English: I imagine that in colloquial and literary English there would be a predominance of French, whereas in science, direct Latin imports would dominate. This may be the object of a future study. It might also be interesting to study whether enhanced precision in other non-Neo-Latin languages requires the use of Latin or do these other languages have “native” words that can replace Latin? As far as German is concerned the number of Latin words is substantially less than in English. However, to understand the Latin influence over German, one must also remember that a very large number of “precision” German words were constructed as Germanic translations of Latin words. These are extremely numerous and represent the invisible transfer of the Latin culture into the German language.<sup>8</sup> However, a quantitative study of this is beyond the scope of this study, which centered on English, because of its essential role as the lingua franca of contemporary science.

Finally, the question proposed as the title of this article has a very definite and simple answer. It is possible to write English extensively, albeit imprecisely, using few or none of the borrowed words. The Christian Lord's prayer is a fine example: it contains 48 words, of which only two are of Latin origin: the French "deliver us" can be easily replaced by the Germanic "free us"; "temptation" is a little more difficult: you would have to go all the way back to Old English to find a good equivalent Germanic word: "costnung"; unfortunately, nobody except Old English scholars would know that "costnung" means "temptation".

Inversely, as noted above, it is virtually impossible to write a proper phrase in English using only borrowed words. The core of any language is contained in its crucial sentence-forming words (prepositions, conjunctions, articles, auxiliary verbs), plus all of the irregular verbs and the most common regular ones. All of this comes from the original Anglo-Saxon base. Thus, English is definitely not a Latin, nor a French language in disguise: it is a Germanic language, of the Anglian sub-family.

To conclude, a word about the relevance of this study in terms of how-to-write a good English medical text. The following is especially true if you are a native speaker of any of the Neo-Latin languages. Some points are essential: (a) roughly 50% of your finished text will be of Latin origin and will consequently contain true cognates to your native language; (b) roughly 30% will be the crucial all-Germanic sentence-forming words (prepositions, conjunctions, articles, auxiliary verbs); prepositions are complicated and you must work hard to avoid mistakes; the other categories behave in a manner similar to your own native speak; (c) roughly 20% will be meaningful Anglo-Saxon words (all the irregular and most of the regular verbs, pronouns, nouns, adjectives, adverbs): irregular verbs must be learnt by heart, but all else poses few or no problems.

## ■ CONFLICT OF INTEREST

The author reports no conflict of interest regarding this study.

## ESCREVER CIÊNCIA COM QUALIDADE: SERIA O INGLÊS CIENTÍFICO UMA LÍNGUA LATINA DISFARÇADA?

**CONTEXTO:** O inglês é a língua franca da ciência; é a língua das duas mais recentes superpotências mundiais e a língua de quatro dos dez maiores produtores de ciência do mundo; é uma língua bastante simples e o idioma mais hibridizado da história, com o latim e o francês contribuindo com aproximadamente 60% do léxico inglês.

O objetivo deste estudo é determinar se a frequência de uso de palavras importadas é uma função do gênero literário.

**MÉTODO:** os textos foram selecionados aleatoriamente de (a) artigos científicos médicos, (b) relatórios financeiros dos jornais, (c) reportagens desportivas, (d) textos literários, (e) inglês coloquial; Para comparação, uma coleção de textos distribuídos de forma semelhante foi selecionada a partir do português; a frequência de ocorrência de palavras latinas ou neolatinas foi determinada nos textos em inglês e na ocorrência de palavras não latinas ou não neolatinas nos textos portugueses; uma análise de variância unidirecional foi utilizada para determinar se diferenças significativas ocorreram entre gêneros nas duas línguas.

**RESULTADOS:** A frequência de ocorrência de palavras latinas / francesas em textos ingleses foi significativamente dependente do gênero literário, sendo máxima em textos científicos médicos e mínima em inglês coloquial; em contraste, a frequência de ocorrência de palavras não latinas em português foi constante ao longo dos mesmos gêneros literários.

**CONCLUSÃO:** O uso de palavras de origem latina ou francesa em inglês é diretamente proporcional à complexidade do gênero literário, fenômeno não observado em português, uma língua neolatina típica.

**PALAVRAS CHAVE:** Educação Médica, Linguagem Científica, Etimologia.

**See also article:** A comment on "Writing Good English: Is scientific English a Latin Language in Disguise?". Available from: <http://www.medicaexpress.net.br/details/321/a-comment-on--writing-good-english--is-scientific-english-a-latin-language-in-disguise-->

## ■ REFERENCES

1. Scimago Journal and country rank: [www.scimagojr.com](http://www.scimagojr.com), assessed on November 15, 2017, 2017 in <http://scimagojr.com/countryrank.php?year=2016&order=itp&ord=desc>
2. Finkenstaedt T, Dieter W. Ordered profusion; studies in dictionaries and the English lexicon. C. Winter. Heidelberg. 1973. ISBN 3-533-02253-6
3. Williams JM. Origins of the English Language. A social and linguistic history. 1986. Free Press. ISBN 0029344700
4. Cioranescu A. Diccionario etimologico rumano. 1966 Gelos, Madrid.
5. Chaucer J. Canterbury Tales; the prologue circa 1400. Translated into modern English by the author.
6. SciELO, Scientific Electronic Library Online: [www.scielo.br](http://www.scielo.br), assessed on November 10 - 20
7. Engel J. Assessed on December 26, 2017 in <https://www.quora.com/If-60-of-English-vocabulary-is-Latin-based-why-is-it-considered-a-Germanic-language>
8. Radici R. Assessed December 26, 2017 in 1. <https://www.quora.com/How-many-words-in-German-borrowed-from-Latin>