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.
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 her5
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 14th 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 Wlliam the Conqueror to
the end of the 14th 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 ca-
tegories: 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
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 Zimmernann 1953), Mighty Aphrodite (Woody Allen, 1995), Pillow Talk (Michael Gordon, 1959), Some like it hot (Billy
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. 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. 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.
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 géneros nas duas línguas.

RESULTADOS: A frequência de ocorrência de palavras latinas / francês 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 géneros 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, Étimologia.

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

REFERENCES

5. Chaucer J. Canterbury Tales; the prologue circa 1400. Translated into modern English by the author.