
A CORPUS-BASED APPROACH TO ANGLICISMS IN THE ROMANIAN ECONOMIC PRESS

GREAVU Arina

Lucian Blaga University of Sibiu, Romania

Abstract:

The paper conducts a quantitative analysis into the phenomenon of English borrowing in Romanian, as reflected in one of the leading economic publications over a period of one year (Capital 2005). Starting from a definition of Anglicisms that is based mainly on formal criteria, borrowed words in the above newspaper are identified and classified according to frequency of occurrence, both in relative and in absolute terms. Thus, Anglicism types and tokens are quantified as regards the proportion they hold in the total number of words in the publication, and in terms of their repetition rate. Special attention is devoted to borrowed words in the lowest and highest frequency ranges, with some of the 100 most frequently used Anglicisms being discussed as regards their necessity in the language.

Keywords: Anglicism, borrowing, word type, word token

1. Introduction and literature review

When Romanian-English contact phenomena are discussed, a very often used term is that of Anglicism. Although this term has gained considerable currency when describing the importation of English words into various European languages (Gorlach 2002), thus occupying a special place within the large spectrum of language contact terminology, it is far from having a unanimously accepted definition in the literature. This section reviews some of the most important definitions given to the term, and sets the conceptual boundaries within which it will be used in the rest of the paper.

In one of the most comprehensive and insightful contributions on the topic to date, Mioara Avram defines an Anglicism as

o unitate lingvistică (nu numai cuvânt, ci și formant, expresie frazeologică, sens sau construcție gramaticală) și chiar tip de pronunțare sau/și de scriere (inclusiv de punctuație) de origine engleză, indiferent de varietatea teritorială a englezei, nu doar din cea britanică. (1997: 11)

Other writers (Dimitrescu 1995, Constantinescu, Popovici and Ștefănescu 2002) operate with a similar classification of Anglicisms, using the term very broadly to cover all categories of borrowed words, regardless of their degree of integration into

the language. Thus, Anglicisms are described as including the whole range of lexical phenomena “from the stage of unadapted neologisms, through that of incompletely adapted words, to that of totally adapted items” (Constantinescu et.al. 2002: 187), and receive the following classification:

a. totally unadapted and not felt to be part of Romanian. These include foreign material ranging from single words (*board, building, challenge, establishment, pet, pub, scholarship*) to multiword stretches of language such as quotation words and code-switching (*to be or not to be, time is money, take it easy, no comment, etc*).

b. words still looking foreign in form. These are more frequently used and felt as more “necessary” than those in the first category, being therefore regarded as more stable and less influenced by the pragmatic factors of communication. Examples include *show biz, exit poll, melting pot, politically correct, baby-sitter, boom, dealer, empowerment, full-time, item, job, know how, public relations, self made man, etc*

c. fully integrated items. These “are no longer felt as Anglicisms or have lost the status of “traveling words” (2002: 188). Examples include *yancheu, troleu, hent, lider, meci, miting, pasa, snec, sut, etc*. In addition to these classes of lexical borrowings, the term Anglicism is also taken to include calques (*brainwashing < spalarea creierelor, cyberspace < ciberspatiu, self-service < autoservire, skinhead < capete rase, skyscraper < zgarie-nori*), semantic loans (*a aplica* added the English meaning of ‘to request’, *a realiza* added the English meaning of ‘to realize’), and pseudo-loans (*tenisman*).

Taking as a reference point the definition given to the term by MDN (2005) as “cuvânt, expresie proprie limbii engleze; cuvânt patruns dintr-o altă limbă și încă neintegrat în aceasta”, Stoichițoiu-Ichim (2006) defines Anglicisms as “*împrumuturi din engleza britanică și americană aflate în curs de adaptare la sistemul limbii române*”, thus excluding completely assimilated loanwords as well as completely unadapted ones. While narrower in its scope as it excludes the classical established borrowings from the class of Anglicisms, this definition includes cases of codeswitches or “borrowed phrases”, and it suggests that a word’s form might be a useful criterion in establishing its status in the lexicon.

Other interpretations of the term give paramount importance to a word’s origin at the expense of formal elements. For example, Filipović (1990:17, quoted in Manolescu 1999) defines Anglicisms in the following way:

every word borrowed from English denoting any thing, idea or concept that belongs to the English civilization; it need not be of English extraction but it must have been adapted in English and integrated in the English vocabulary.

While it indicates as Anglicisms all words denoting English realities, even when another language served as a vehicle for their transmission, this definition also includes words of exotic origin such as *penguin, banjo, rodeo, mustang, chinook* or *toboggan* due to the fact that “they have been made known and spread by English.” Several Romanian researchers have adopted and used this approach in their attempt to establish the extent of the English element in the Romanian language. Thus, based on this acceptance of the term, Manolescu (1999) and Ciobanu (1996) compiled lists of

Anglicisms in Romanian, with the result that these lists include words of English origin whose transmission into Romanian was mediated by a third language (usually French, but also German, Italian or Russian), even when the origin indicated in the dictionary is not English. For example, Ciobanu's inventory contains words such as *biftec*, *galon*, *Anglican*, *Newton*, *raglan*, words on which the French intermediacy left unmistakable traces, making Englishness difficult to recognize and claim.

Since none of the principles of Anglicism identification presented above are unanimously accepted by language contact researchers and lexicographers alike, dictionaries are not always reliable instruments in establishing the etymology of a word, different dictionaries sometimes indicating different etymologies for the same word. Some dictionaries mark only the last etymological link and not the original source, so that words that were not borrowed straight from English, but transmitted via another language (French, German or Russian) are presented without any reference to English. Other dictionaries (Marele dictionar de neologisme 2006, Micul dictionar academic 2001, Dictionarul explicativ al limbii romane 1998) mark all etymological links, but without necessarily indicating their chronological order. Still others (DOOM 2005) would indicate English origin for words that are assigned multiple etymology in other lexicographic works. To give just a few examples, words in our corpus described as both 'fr.' and 'engl.' in MDN 2006 and MDA 2001, are shown to be Anglicisms in DOOM 2005: *brainstorming*, *bridge*, *clearing*, *cocktail*, *derby*, *designer*, *driver*, *dumping*, *grill*, *gadget*, *hamburger*, *hobby*, *marketing*, *planning*, *rap*, *shopping*, *stray*, *staff*, *striptease*. Such inconsistencies make statistical counts of Anglicisms based on dictionaries extremely difficult if not impossible, a situation remarked by several authors in the field (Bantaş 1982, Avram 1997, Constantinescu et al 2002).

The lack of terminological agreement regarding the definition of Anglicisms is also reflected in the discrepancies existing between various quantitative studies on English loanwords in Romanian. For example, in a 1982 article based on the study of bilingual dictionaries at that time, Bantaş reports on the existence of nearly four thousand English words. A count by Ciobanu (1996), on the other hand, based on the main dictionaries published in the last four decades of the 20th century, newspapers, magazines, almanacs, articles as well as oral sources, reveals a corpus of only 1,400 Anglicisms. Her inventory includes simple words, but also set phrases and phraseological units, while excluding words belonging to highly specialized terminologies. A list similar to that of Ciobanu's, but based on a broader time span in Romanian/ English contact was subsequently drawn up by Manolescu (1999), who examined lexicographic works covering a vast period of time, from the first modern dictionary published in 1898 and continuing up to the mid-1990s, together with the vocabulary of some newspapers from the last decade. The 3,820 Anglicisms listed include acronyms, pseudo-Anglicisms and semantic loans together with single word borrowings - all of them representing 'the English element in contemporary Romanian'.

The perspective adopted in the definition of the term Anglicism in the present study is a synchronic and formal one. Thus, we consider a word's form to be of paramount importance in separating it from native vocabulary. From this perspective,

an Anglicism is defined as any English lexical element in the economic publication *Capital* that can be *formally* related to English. This acceptance of the term is in line with the definition it was given by other researchers (Stoichițoiu 2006, Gorlach 2002, Onysko 2007), who regard as an Anglicism any word “recognized in its form (spelling, pronunciation, morphology) as coming from English” (Onysko, 2007: 90), and it also captures the essence of the definition of borrowing as present in several influential works on the topic. Thus, Haugen (1950), Thomason and Kaufman (1988), and Myers-Scotton (1992, 1993) among others, define linguistic borrowing in a way that suggests the idea of direct contact between the source and the recipient language, without the intermediacy of other languages: borrowing is the deliberate reproduction or incorporation of elements found in one language in another one, as a result of specific socio-economic conditions surrounding the contact situation.

This definition of Anglicisms excludes integrated loanwords, which are English borrowings only in a historic sense, leaving for analysis only those recent English elements which have not been yet adapted to Romanian. In addition to this, the present study will be concerned only with simple, one-word Anglicisms, its main objectives being to measure the numerical impact of these words in a corpus of journalistic texts, to describe their distribution across different frequency ranges, and finally to identify the most often used and therefore most stable English words in this corpus.

2. Methodology of research

The source of the corpus was the economic publication *Capital* on CD-Rom, consisting of Adobe PDF files. This raw data underwent a series of processing procedures, i.e. Optical Character Recognition, sentence splitting, tokenization and part-of-speech tagging and lemmatization. The texts thus obtained, amounting to 78,067 types and 2,891,880 tokens were tapped using dedicated software tools designed specifically for this project. Following a process of Anglicism identification from decontextualized word lists and subsequent filtering so as to eliminate Romanian homographs and proper names, the final results show a number of 2,135 types and 27,928 tokens, amounting to 2.73% of all types and 0.96% of all tokens in the corpus. This represents an average of one Anglicism for every 36.56 word types and every 103.54 tokens. The discrepancy existing between the numerical proportion Anglicisms hold in terms of types and their representation as far as occurrences are concerned indicates a low repetition rate for borrowed words. This, in turn, can be a result of the novel character these words have in the language, and of their low level of acceptability, or can simply reflect the reduced incidence in discourse of the concepts they are used to designate.

However, since the occurrence of many of the Anglicisms identified at this stage is restricted to certain phrases and compounds representing direct importations from English, phrasal units and simple borrowings had to be counted separately. This

paper investigates only the use of the second category, calculations being standardized for one million words.

3. Presentation and discussion of results

Following the process of English words identification, filtering and categorization, simple Anglicisms were analysed in terms of their distribution across different frequency ranges. The results of this analysis are presented in Figure 1 below. Thus, as many as 50% of all borrowings are used merely once in the standardized corpus, 25% have a token frequency of 2, while 5% appear three or four times in our data. At the opposite pole, only 9% of all English words are used more than 10 times, while a very low percentage of 3% appear more than 50 times and can therefore be considered as being well established in Romanian.

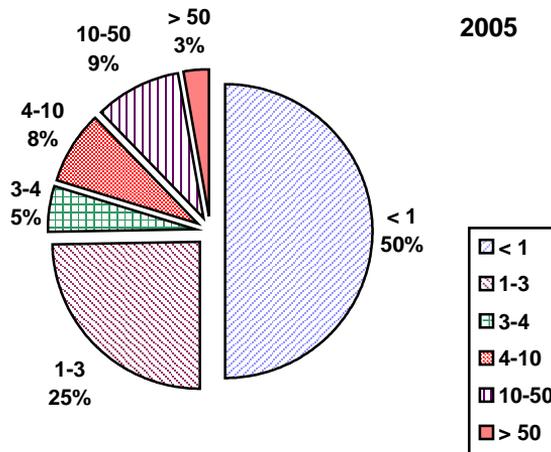


Figure 1. Distribution of simple Anglicisms according to token frequency in Capital 2005 (rounded values)

Anglicisms used merely three times or less include words such as those underlined in the following phrases: *binemeritat loc 3 worldwide, strategie winwin, program de trainee, standing financiar, credite overnight, echipamente outdoors, magazine multibrand, piata hospitality, tehnici de fundraising comunitar, conceptul de empowerment, piata e-commerce, cluster IT, modulul de back-office, contracte de wholesale, servicii upstream, situatia de sold-out, fonduri private de equity, carrieri internationale, derivate de tipul call si put, procedura write-off, credit de tip revolving, etc. Such rarely used borrowings have received different names in the language contract literature- nonce loans (Poplack 1980), spontaneous borrowings or speech borrowings (Grosjean 2001)- being inherent elements in the speech of some bilinguals, rather than a matter of the language as a whole.*

From a quantitative point of view, it is particularly important to see which of the Anglicisms used in the corpus of *Capital 2005* are most often used, and therefore showing better chances of being adopted in the language, or in the speech of monolinguals. This is why a list of the 100 most frequently used simple Anglicisms in the studied corpus was drawn up, showing the token frequency (TF) or number of occurrences for each Anglicism lemma. This ranking according to frequency of occurrence is given below.

Nr.	Lemma	TF	Nr.	Lemma	TF	Nr.	Lemma	TF
1	marketing	858	35	ATM	97	69	server	54
2	USD	835	36	accesa	95	70	wireless	54
3	management	798	37	mass media	94	71	leadership	53
4	manager	729	38	notebook	86	72	multimedia	50
5	Fax	702	39	accesare	84	73	rebranding	50
6	internet	694	40	PR	84	74	subsidiara	49
7	card	648	41	mall	83	75	outsourcing	48
8	trend	589	42	soft	83	76	SUV	48
9	email	572	43	CD	81	77	chart	46
10	Site	557	44	futures	81	78	trader	46
11	leasing	522	45	web	80	79	Gprs	45
12	brand	420	46	CV	77	80	player	45
13	IT	397	47	laptop	76	81	lista	44
14	business	360	48	PDA	75	82	wi fi	43
15	retail	350	49	second hand	75	83	brokeraj	42
16	online	315	50	Pet	74	84	cash	42
17	software	299	51	fitness	71	85	GSM	42
18	service	265	52	Ipod	71	86	hardware	42
19	spot	261	53	low cost	71	87	living	42
20	supermarket	221	54	sponsorizare	71	88	bond	41
21	dealer	214	55	bluetooth	67	89	master	41
22	training	208	56	lobby	67	90	senior	38
23	rating	206	57	IT&C	64	91	showroom	38
24	computer	200	58	subchart	62	92	weekend	38
25	design	184	59	computerizat	60	93	GRP	36
26	broker	150	60	discount	59	94	hobby	36
27	PC	134	61	factoring	59	95	advertiser	34
28	hipermarket	132	62	Forex	59	96	know how	34
29	managerial	124	63	charter	57	97	coach	33
30	DVD	121	64	accesat	55	98	flash	33
31	retailer	121	65	fast food	55	99	masterat	33
32	Job	119	66	premium	55	100	print	33
33	bonus	116	67	SMS	55			
34	gadget	105	68	boom	54			

Table 1. The 100 most frequent Anglicisms in Capital 2005

This list shows that many of the words that have a very high frequency of occurrence are used to designate concepts, objects or activities from the field of economy as well as from other related fields, e.g. information technology, entertainment. These Anglicisms are prototypical examples of cultural loans, and their adoption into the language can be regarded as being a necessity, i.e. they have an instrumental role in filling lexical gaps in the vocabulary of Romanian. Such loans appear as a result of an attempted cognitive simplification on the speaker's part, who tries to find the simplest linguistic strategy in naming new things, as it is easier and more economical to use ready-made designations than to coin new words. Examples in our corpus include *internet*, *Bluetooth*, *boom*, *email*, *site*, *leasing*, *rating*, *broker*, *charter*, *factoring*, *fast-food*, *charter*, *IT*, *outsourcing*, *training* to name just a few, as well as *marketing*, *management* and *manager* – words which have already been adopted in the language, being actually recorded by older dictionaries (DEX 1975).

However, other frequently used words have perfectly acceptable synonyms in Romanian. Although they have a cultural content too, these words can be described as redundant, as they do not fill any lexical gaps in the language but rather double already existing words. They have a symbolic rather than a referential value, being indexical of a higher social status, fashion and prestige usually associated with English as the lingua franca of the contemporary world. Such borrowings are known in the literature as gratuitous or unnecessary loans, and are usually condemned by language purists as reflecting a poor knowledge of one's native language, linguistic snobbery or simply laziness, haste and the law of least effort. This is the case with *job* (*slujba*, *loc de munca*), *trend* (*tendinta*, *evolutie*), *brand* (*marca*), *business* (*afacere*, *firma*, *companie*), *cash* (*numerar*), *bond* (*obligatiune*), and many others.

Even though the dichotomy between necessary and unnecessary loans is generally accepted in the language contact literature, some writers have recently challenged it (Backus 1996, Onysko 2007), arguing that any borrowed word can be considered as filling a lexical gap, mainly because translation equivalents are not exact equivalents, the encyclopedic knowledge a speaker has about a certain word seldom matching the connotations rendered by the resources of another language. From this perspective, it can be maintained that the English words listed above carry slightly different meanings from those of their Romanian synonyms, and can be expected to undergo various processes of semantic change, e.g. meaning narrowing, meaning broadening, which will further separate them from native words.

4. Conclusions

The quantitative analysis of Anglicisms in the studied corpus has shown that only a relatively small proportion of these words have a high frequency of occurrence, more than 50% of all English words being used merely once or twice in our data. Such a limited usage could reflect the new character these words have in the language, or merely the fact that the concepts they are used to designate were not very often discussed in the studied corpus. The relatively large proportion borrowed words have

in the total number of words in the corpus seems to testify to an increasing intensity of contact between English and Romanian, and may point towards an upward trend in the phenomenon of borrowing between the two languages. Although at a more detailed level of analysis, the borrowed words can be broadly divided into necessary and unnecessary loans, this distinction is empirically weak as the line separating the two classes remains arbitrary and therefore difficult to draw.

5. Acknowledgements

The corpus processing tasks were performed by Eckhard Bick (researcher) and Tino Didriksen (student assistant), from the Institute for Language and Communication (ISK) at the University of Southern Denmark. The tagging was done using the MSD tagger developed by the Research Institute for Artificial Intelligence of The Romanian Academy, under Professor Dan Tufiş' supervision. The pos-tagged corpus is available at <http://corp.hum.sdu.dk/cqp.ro.html>

6. References

- Avram, M. (1997), *Anglicismele în limba română actuală*. Conferință prezentată la Academia Română. București: Editura Academiei Române.
- Bantaș, A. (1982), Aspects of applied semantics: for modernizing bilingual dictionaries. *Revue Roumaine de Linguistique* XXVII, 3: 219-226.
- CD Capital2005*. București: Ringier România.
- Ciobanu, G. (1996), *Anglicismele în limba română*. Timișoara: Amphora.
- Constantinescu, I., Popoviciv V., and Ștefănescu, A. (2002), Romanian. In Görlach, Manfred (ed.), pp. 168-194.
- Dicționar ortografic, ortoepic și morfologic al limbii române, ediția a II-a*, (2005) Academia Română, Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti”, București: Editura Univers Enciclopedic.
- Dicționarul explicativ al limbii române*, (1975) Academia Română, Institutul de Lingvistică din București: Editura Academiei.
- Dicționarul explicativ al limbii române*, (1998), Academia Română, Institutul de Lingvistică „Iorgu Iordan”, București: Editura Univers Enciclopedic.
- Dimitrescu, F. (1995), *Dinamica lexicului românesc*. Cluj-Napoca: Editura Clusium.
- Görlach, M. (ed.) (2002), *English in Europe*. New York: Oxford University Press.
- Grosjean, F. (2001), *Life with Two Languages: An Introduction to Bilingualism*. Cambridge, Massachusetts, London: Harvard University Press.
- Manolescu, Z. (1999), *The English Element in Contemporary Romanian*. București: Conspress.
- Marcu, F. (2006), *Marele dicționar de neologisme*. București: Editura Saeculum.
- Micul dicționar academic*, (2001), Academia Română, Institutul de Lingvistică „Iorgu Iordan”, București: Editura Univers Enciclopedic.
- Myers Scotton, C. (1993), *Duelling Languages: Grammatical Structure in Codeswitching*. New York: Oxford University Press.

- Onysko, A. (2007), *Anglicisms in German: Borrowing, Lexical Productivity, and Written codeswitching*. Berlin, New York: Walter de Gruyter.
- Stoichițoiu-Ichim, A. (2006), *Aspecte ale influenței engleze în româna actuală*. București: Editura Universității din București.
- Thomason, S.G., Kaufman, T. (1988), *Language Contact, Creolization, and Genetic Linguistics*. Berkeley, Los Angeles, London: University of California Press.