

BARE ENGLISH-ORIGIN NOUN-PHRASES IN ROMANIAN

Arina GREAVU

Lucian Blaga University, Sibiu

Abstract: English words borrowed in Romanian are most often adapted so as to conform to the morphosyntactic rules of the recipient language. However, sometimes borrowings remain uninflected for Romanian grammatical categories, by not receiving the relevant suffixes for gender, number, case, and definiteness. These unintegrated forms are known in the literature as bare forms. The present paper analyses such cases of English-origin bare nouns and noun phrases in Romanian, and tries to explain the structural, social and functional mechanisms that trigger their appearance.

Keywords: borrowing, Anglicism, integration, bare form

1. Introduction and definition of concepts

Words borrowed from one language into another one usually undergo a process of integration into the morphosyntactic structures of the recipient language. This integration process is governed by a number of linguistic, social and psychological factors which combine according to the specific circumstances surrounding the contact situation. The factors most often described in the literature are: the word class of the borrowed words and the typology of the two languages in contact, the frequency with which these words are used and their age of existence in the recipient language, the role they fulfil in this language (to fill lexical gaps or to double already existing words), the bilingual ability of the recipient language speakers, their loyalty to their native tongue and attitude towards borrowing in general. However, none of these factors can be expected to act categorically in the integration process, their complex and sometimes highly idiosyncratic interplay rendering any deterministic approach to the study of linguistic integration in contact situations untenable.

In this context, a problem which has attracted the attention of researchers studying both borrowing and codeswitching phenomena, is that of foreign elements which, for some reason, escape the adaptation routines imposed by the recipient language. These forms are known in the literature as bare forms (Myers-Scotton 1993, Backus 1996, Boumans 1998) and have been defined as “EL morphemes without inflections, or modifying function words from either language.” (Myers-Scotton, 1993: 112), where EL is the embedded language (also known as the source or donor language) while ML stands for the matrix language (the recipient or borrowing language). In other words, bare forms are foreign words which do not meet the structural requirements of the host language, by not being marked, either morphologically or syntactically, for the various grammatical categories of this language.

This paper attempts to analyse bare English-origin nouns and noun phrases in a corpus of contemporary Romanian journalistic prose. The term *borrowing* will be used as defined by Haugen, i.e. “the attempted reproduction in one language of patterns previously found in another” (1950: 212), and by other researchers studying English

borrowings or Anglicisms in various European languages. For example, Gorlach regards as an Anglicism any word “if it is recognizably English in form (spelling, pronunciation, morphology)” (2005: xviii), and Onysko applies the term to “any instance of an English lexical, structural, and phonological element (...) that can be formally related to English” (2007: 90).

Partly following these writers, we are going to include in the present analysis English-origin elements that show *formal identity* with their source-language models. This methodological approach excludes established loanwords (*chec, șut, miting*, etc) from the scope of the study, focusing on recent and obtrusive importations from English which have preserved their foreign graphemic character. Such importations include simple words and compounds (e.g. *brand, deal, job, brainstorming, notebook, offshore*) as well as longer phrases (e.g. *call center, day spa, managing director, brand awareness, business intelligence, chief executive officer*). The terms *English element* will also be used generically to cover both simple and phrasal Anglicisms.

2. Corpusⁱ and methodology of research

The source of the corpus consisted of the business magazine *Capital* on CD-Rom (the 2005 issues), 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 lemmatizationⁱⁱ. The texts thus obtained, amounting to 2,891,880 tokens, were tapped using customized software tools designed specifically for this project. The first stage of this process was the generation of decontextualized word lists showing all the individual word types in the corpus, and thus facilitating a faster identification of possible English words. According to the definition given to the term borrowing in the previous section, a number of 4,495 word types and 63,175 word tokens were elicited from a corpus of 78,068 types. Following a filtering process which eliminated Romanian homographs (e.g. *deal, fast, brand*) and proper nouns, we arrived at the final amount of individual borrowings in the *Capital* 2005 corpus. The next stage consisted of the identification of phrasal constructions in the total of English elements, and the separation of the two classes (simple vs phrasal Anglicisms) based on the criterion of length. The results of this stage of the analysis are presented in the table below.

	Types	Tokens	Frequency (token/type)
Total nr. of words	78,067	2,891,880	37.07
Total nr. of Anglicisms	2,135	27,928	13.09
Nr. of simple Anglicisms	1,442	20,534	13.99
Nr. of phrasal Anglicisms	860	2,497	2.91

Table 1. Anglicisms in *Capital* 2005 by structural type

Next, all English words and phrases found in the corpus were tagged for part of speech and grammatical categories. As a result of this procedure we identified a number of 1,296 English-origin nouns (including compounds) and 773 English-origin noun phrases (made up of two or more words). This statistic is presented in the table below:

	Types	Tokens	Frequency (token/type)
Nouns	1,296	19,036	14.68
Noun phrases	773	2,317	2.99

Table 2. Number of nouns and noun phrases in *Capital 2005*

Once the tagged corpus of simple and phrasal Anglicisms was in place, it was possible to draw some conclusions regarding the representation of bare forms in the total of borrowings. In the present paper, the sole focus on attention will be on the grammatical category of number, as we will attempt to identify those English elements which have plural referents, but are used without the obligatory plural morphemes required by Romanian norm.

3. Presentation of results

In Romanian, nouns can be divided into countable and uncountable. Generally, nouns designating concrete animate and inanimate entities are countable (e.g. *scaun, copac, copil, familie*), while nouns designating mass or abstract concepts are uncountable (e.g. *aur, miere, lapte*). Within the class of countable nouns, a special subclass is constituted by invariable plurals (e.g. *arici, ochi, unchi, învățătoare, nume*), which show formal identity in the singular and in the plural. The grammatical category of number is marked morphologically on most nouns by a system of rich inflections, and by a number of intra-word vowel and consonant alternations. Number is also marked syntactically on determiners such as adjectives and pronouns, which must agree grammatically with the head they modify (Guțu-Romalo, 2005: 69-70).

From the perspective of these general rules, several cases of bare English-origin nouns and noun phrases have been identified in the *Capital 2005* corpus. A first case is represented by those countable nouns designating concrete concepts (products, professions, events, institutions), which remain uninflected for plural although clearly having a plural referent. Many of these words appear inside prepositional phrases headed by *de*, and functioning as modifiers of nominal heads derived from transitive verbs (e.g. *vânzare, producție, utilizare*). Although plural marking is not categorical in *de* phrases, the singular nouns that participate in these constructions either belong to the class of mass, abstract nouns or have a prototypically singular referent (e.g. *vânzare de carte, producător de film, drum de țară, casă de vacanță, ținută de seară, problemă de timp, iubire de mamă, apă de izvor*). This is why we believe that the English words and phrases in the sentences below, as well as other similar cases, can be regarded as morphologically bare forms:

- (1) (...) traficul de **SMS** a crescut cu 90% față de aceeași perioadă a anului trecut. (...). De asemenea, și numărul de utilizatori de **SMS** este mai mare cu 50%.
- (2) Spectacol pe piața de **notebook**. Vânzările au crescut cu 60% în 2004 față de 2003.
- (3) Vânzările de **PDA** în lume (headline)
- (4) Cât despre piața de **Smartphone**, aceasta a înregistrat o creștere de 135% în raport cu perioada similară a anului trecut.
- (5) Am fost prima generație de **internet marketing specialist** din Europa.
- (6) (...) numărul stagiatorilor cuprinși în noul **program de "trainee"**.
- (7) Vrem să inițiem și cursuri de formare de **coach**.
- (8) Noul lanț de **"fashion café"** va avea 12 locații (...)

The obligatory nature of the plural morpheme in the sentences above is emphasized by its actual realization in similar contexts, the studied corpus containing constructions such as *număr de SMS-uri, vânzările de PDA-uri, piața de PDA-uri, vânzările de notebook-uri, producătorii de notebookuri, flota de smartphonuri si PDA-uri, număr de GRP-uri, etc.* Moreover, similar bare words and phrases are sometimes part of coordinated constructions in which the other elements (mostly native words or more established borrowings) are marked for plural. In this context, we believe that the English-origin lexical items in the examples below should follow the same specifications for number as their native counterparts:

- (9) Se vând circa 25.000 de bucăți pe an, incluzând **PC desktop, notebook și servere**.
- (10) Alte surse de interferență sunt *telefoanele fără fir, cuptoarele cu microunde, walkie-talkie și alte rețele fără fir vecine*.
- (11) (...) a crescut numărul promoționalelor scumpe din metale prețioase (...) *portțigarete, stilouri, port-card-visit, moneyclip - sau din lemn prețios lăcuit, (...)*
- (12) În vest, lanțurile de *cafenele și fast food*, cum ar fi Starbucks, (...), au făcut businessmenii să coboare din birouri.
- (13) (...) hotelierii spun că (...) dacă n-ar fi susținută de turismul de afaceri (*conferințe, seminare, teambuilding*), treaba ar merge mult mai prost.

Another category of bare elements is constituted by heads of phrases used in the singular, although their modifiers are clearly marked for plural. As the agreement in number between a head and its modifier is categorical in Romanian, we believe that the sentences below contain the most obvious examples of bare English-origin nominals in the studied corpus:

- (14) Probabil că multora dintre noi (...), gândul ne zboară la *una dintre cele mai celebre businesswoman* (...).
- (15) Numărul de **call center locale** este în creștere, iar afacerea a început să înflorească abia anul trecut.
- (16) Primul televizor din lume cu ecran OLED de *40 de inch*. Samsung Electronics a creat primul ecran OLED (...) cu diagonala de *40 de inci,*

Other words and phrases borrowed from English and lacking Romanian plural morphology are used as subject predicatives, a syntactic position which involves the agreement with the subject of the sentence:

- (17) *TINERII NOȘTRI SUNT "JOBHOPPER"*. (headline)
- (18) Da, se vede în numărul de *oameni* care se pregătesc pentru a deveni **coach**, în creștere cu 30,50% la sută pe an.

However, since the requirement for plurality is categorical in such cases, its disregard constituting a clear divergence from the grammatical rules of Romanian, in very many cases when English borrowings enter such "flawed" constructions they are "flagged", thus being dislocated from the main morphosyntactic frame of the sentence and escaping the integration process. A flag is a strategy the writer uses in order to signal a hesitation or "an actual change in language so that morphosyntactic integration is not necessary anymore" (Backus 1996: 167). Flagging can be accomplished by placing the word in question between inverted commas, between brackets, or by using a metalinguistic commentary to introduce it (e.g. *așa-numitele/așa-numiții, așa-zisele/ așa-ziișii, numite/ numiți*). The sentences below illustrate the occurrence of bare nominals which are flagged by one or several of these strategies:

- (19) Să nu uităm că virușii care au avut cel mai mult succes anul trecut au fost *așa-numiții mass-mailer*.
- (20) Faimoasele mici dispozitive de stocare *numite "pen drive"* au înlocuit aproape complet dischetele
- (21) (...) mi-au atras atenția *așa zisele "quiet room"*, niște birouri cu un perete din sticlă, (...).
- (22) În prezent, în Statele Unite, cele mai numeroase sunt *așa-numitele "day spa"*, centre amplasate în marile orașe, (...).
- (23) (...) un număr important de filiale ale unor rețele multinaționale au drept acționari firme înregistrate în paradise fiscale - *așa-numitele off-shore*.
- (24) (...) nevoile clientului, care pot varia de la o radiografie sumară asupra pietei într-un anumit moment, până la *analize în profunzime ("consumer insight")*.
- (25) Acest lucru a dus la sporirea numărului de *pliante promoționale (leaflet)* editate de companiile de retail.

On the basis of the criteria discussed above (singular nouns with plural referents inserted into *de* phrases, flawed parallelism with coordinated plural nouns, flawed agreement between singular heads and plural modifiers) we have arrived at the following statistics regarding the representation of bare nominals in the total of English-origin importations with plural referents. Tables 3 and 4 below show the distribution of various plural formation patterns across the separate classes of simple and phrasal Anglicisms. The independent analysis of these two classes reveals the fundamental differences existing between them as regards their integration into the morphosyntactic structures of Romanian.

Plural formation	Types	%	Tokens	%	Frequency (tokens/types)
Romanian plural	453	90.96%	3,296	94.73%	7.27
English plural	22	4.41%	129	3.70%	5.86
Double plural	4	0.80%	9	0.25%	2.20
Bare plural	19	3.81%	42	1.20%	2.21
Total nouns with plural referents	498	100%	3,479	100%	6.98

Table 3. Plural formation in simple Anglicism (*Capital* 2005)

Plural formation	Types	%	Tokens	%	Frequency (tokens/types)
Romanian plural	32	27.77%	53	26.38%	1.65
English plural	66	60.18%	106	52.73%	1.60
Zero plural	14	12.03%	42	20.89%	3.00
Total NPs with plural referents	112	100%	201	100%	1.7

Table 4. Plural formation in phrasal Anglicisms (*Capital* 2005)

The statistical count of bare nouns (types and tokens) shows the very limited scope of this phenomenon. Thus, only 3.81 % of all nominal borrowings having a plural referent, and only 1.20% of their tokens lack plural marking. The figures in Table 3 also indicate a very low repetition rate for bare nouns (slightly above 2), some of which appear merely once in the corpus (e.g. *businesswoman*, *moneyclip*, *jobhopper*). However, English-origin noun phrases which lack Romanian plural morphemes represent a significantly higher percentage of the total of phrases with plural referents,

both in terms of word types and in terms of their number of occurrences. In addition to this, they tend to be repeated more often than their structurally simpler counterparts. We believe that this situation establishes a direct link between the use of bare forms and low frequency of occurrence or limited loanword diffusion across the speech community. The various factors that can be used to explain the employment of bare forms in the studied corpus will be discussed in the following section.

4. Discussion of findings

Bare forms have been described in the language contact literature as “a strategy to avoid clashes between ML and EL morpheme order.” (Myers-Scotton, 1993: 112). In the particular examples discussed in this paper such “clashes” may result from the different number specifications for English and Romanian nominal modifiers. While English most often uses singular nouns in this position (e.g. *coach training, smartphone sales, PDA market, trainee program, SMS traffic, fast-food chain*) Romanian prefers plurals in the analogical construction involving a prepositional phrase (e.g. *formare de instructori, vânzări de telefoane, piața de PDA-uri, program pentru stagiaari, trafic de mesaje, lanț de restaurante*). From this perspective, zero plural marking in borrowed nouns can be interpreted as proving the connection between bilingual ability in the source language and loanword integration: more proficient speakers of English can be expected to show greater interference from this language and preserve some of its structures when speaking Romanian, too, for example by producing bare plurals.

Another explanation for some of the examples presented above (sentences 1 to 9 and 12) could be the employment of the English element on the model of constructions like “vânzare de carte”, “producător de film”, which use singular nouns in order to convey a plural meaning. The existence of such constructions in monolingual discourse offers borrowings “a foot in the door” (Heath 1989, quoted in Myers-Scotton 1993: 154), and explains why the bare English elements entering these phrases are very rarely flagged: since they do not overtly contradict any categorical rule of Romanian grammar, they do not have to be dislocated from the morphosyntactic frame of the sentence, at least not so badly as those borrowings which break agreement rules for example.

In other cases (examples 7 and 16) a phonetic factor might trigger the avoidance of the plural form. Thus, the difficulty of adding a plural morpheme to *-ch* ending masculine nouns has been noted in the literature before, sometimes plural formation leading to the paradoxical and difficult to explain addition of the *-er* suffix on these nouns (Ciobanu, 2004: 139): E. *coach*- pl. *coaches* > R. sg. *coach*- pl. *coacheri*. Another illustration of the way in which this formal constraint can sometimes prevent plural formation of *-ch* ending borrowings is the noun *inch*. In sentence (16), the word appears both as a bare plural and as the integrated, pluralized *inci*, a situation which might suggest the idea that the writer perceives the two forms as being formally and semantically equivalent. However, the suffixed *inchi* does appear for 22 times in the corpus of *Capital* 2005, while the bare form *inch* is used only twice. This proves that the idea of a formal constraint barring plural suffixation of English nouns borrowed in Romanian should not be overemphasized, even if it can be used to explain particular cases.

A phonological reason, too, may underlie the bare form of *walkie-talkie* in sentence (10) and of *trainee* in sentence (6). The pronunciation of these borrowings places them in the inflectional class of Romanian invariable plurals (e.g. *ochi, pui, unchi*), which partly explains the omission of the plural morpheme. The plural formation behaviour of *walkie-talkie, coach, trainee*, and to a lesser extent of *inch* seems to confirm Treffers-Daller's (2001) hypothesis that the perceived equivalence or congruence between the integrated and unintegrated forms in a language can determine integration or, on the contrary, lack of it. Thus, since there is very little or no difference between the singular and the plural form of the words discussed above, the addition of the plural suffix would be not only redundant and thus optional, but also phonetically difficult. These examples also confirm Ciobanu's (2004: 140) observation that a special preference for invariable plurals is shown by *-i* ending singular nouns.

A special situation in our corpus is represented by those English phrases headed by nouns showing formal identity with some native words (examples 5 and 15). A language is very unlikely to borrow a word form which already exists in its stock, in such cases semantic loans being preferred to loanwords as such. For example, since Romanian would not borrow the English *specialist* and *center* outside their larger codeswitching environment, we believe that using a plural morpheme on *internet marketing specialist* in sentence 5 above or on *call center* in sentence 15 may have seemed odd to the writer to purely stylistic and pragmatic reasons. Those phrasal importations from English which show Romanian plural affixation, are headed either by established borrowings (*area sales manageri, middle manageri, DVD playere, etc*) or by bilingual homographs (*business planuri, compact discuri*).

Another factor that has been used in the literature in order to explain the integration of borrowings is their frequency of occurrence in a given corpus. Although frequency is highly dependent on corpus size (a word may be found to be more frequently used in a large than in a small corpus) and on the frequency with which a given concept is discussed (Boumans 1998), the language contact literature provides rich evidence that this factor can impact directly on the integration process. For example, in a study of English loanwords in French, Poplack, Sankoff and Miller (1988) showed that only 85% of the nouns used by a single speaker were integrated morphologically with respect to number, as compared to 98.4% of words used by more than ten speakers. The same correlation was found between frequency and consistency of gender assignment, leading to the conclusion that increased diffusion of a word is strongly correlated with host language affixation.

Low frequency of occurrence seems to be directly linked to the employment of English bare nominals in the *Capital 2005* corpus. Thus, with very few exceptions (*call center* with 23 occurrences and *baby sitter* with 9 occurrences), all the phrasal importations that show zero plural marking have a very low frequency (below 3) in all their forms (both bare and inflected), more than 50% of all these phrases actually occurring only once in the whole corpus (e.g. *consumer insight, day spa, financial disclosure, internet marketing specialist, mass-mailer, quiet room, and passenger car*). However, this situation changes dramatically when we turn to simple nouns: very low frequency words (having three or fewer occurrences both as bare and as inflected forms) represent the exception rather than the rule, merely 5 words out of the total entering this category: *businesswoman* (with 2 occurrences), *jobhopper* (with 1

occurrence), *leaflet* (with 3 occurrences), *moneyclip* (with 1 occurrence) and *trainee* (with 1 occurrence).

Other English-origin nouns occasionally used without plural endings are more widely diffused in the corpus, although they are used more often in their integrated form than as invariable plurals. Thus, *notebook* (87 occurrences) is inflected for plural according to Romanian morphology in 45 instances and uninflected in just one, *off-shore/ offshore* (17 occurrences) uses the Romanian plural suffix in 6 cases and lacks it in two, *PDA* (75 occurrences) has 14 inflected occurrences and three bare ones, *fast-food/ fastfood* (26 occurrences) has ten pluralized occurrences and appears without the plural suffix once, *teambuilding* (56 occurrences) is inflected for plural in 28 instances and uninflected once, *smartphone* (14 occurrences) has three plural occurrence and one bare occurrence, *SMS* (55 occurrences) is inflected for plural in 20 instances and uninflected in 3. These examples reinforce the idea of frequency of occurrence as promoting integration: the more often a borrowed word is used the more inflected it tends to be.

Other bare English-origin nouns which are also used in the plural in our corpus are *desktopuri*, *insighturi*, *café-uri* (*în Internet café-uri*), *leaflet-uri*, *spa-uri*, *baby-sitteri*. Thus, the occurrence of bare forms seems to be variable even at the level of a single word, far from all instances of a borrowing occurring as bare forms. Moreover, as the examples in the previous section indicate, there are several different morphosyntactic contexts in which these forms can occur. From this perspective, it is clear that frequency of occurrence or structural factors on their own cannot satisfactorily explain why the morphological integration of borrowings into the recipient language is avoided in some cases but not in most others. A full account of this phenomenon is possible only if socio- and psycho-linguistic factors are considered.

A factor which has been often used in the literature in order to explain loanword integration is the bilingual ability of the borrowing language speakers, with speakers who are more proficient in the source language tending to show greater interference from this language, and therefore avoid integration more than those who are less proficient (Haugen 1956, Baetens Beardsmore 1982, Myers-Scotton 1993, Backus 1996). For example, in a study of the speech of several generations of Turkish immigrants to the Netherlands, Backus (1996) finds that first generation speakers (consequently those least proficient in the source language) produce the fewest bare Dutch loanwords and EL marked forms. On a related note, speakers from later generations use longer fragments of foreign material (phrases or whole sentences) than those from the first generation.

In our case, it can be argued that those writers who have a good command of English or use it regularly in their activity, will employ more bare forms as well as longer and more complex phrases from this language than Romanian dominant writers. Thus, the poor integration of phrases transferred from English can be explained not only as a consequence of their low frequency of occurrence in the corpus, but also as resulting from the fact that they are used by more proficient speakers who are subject to greater interference from the source language. However, this is only a tentative hypothesis which cannot be verified at this stage of research, as we have no information about the number of writers contributing to the publication or their bilingual ability.

Finally, the integration of words borrowed from one language into another can be influenced by attitudinal factors, such as the borrowing speakers' wish to sound like source language speakers. This attitude towards borrowing can be promoted by certain values attached to foreign sounding or looking words, for example social prestige or fashion, and manifests itself more prominently in the discourse of technology and of other related fields, being indicative of a mixed, modern identity (Pandharipande 1990, Stoichițoiu-Ichim 2001). Given the prominent role English is playing today in the global linguistic environment, we believe that such attitudinal factors can partly account for the morphosyntactic behavior of English words and phrases borrowed in Romanian.

5. Conclusions and outlook

The analysis conducted in this paper has shown that bare English-origin nominals represent a relatively low percentage both in the total number of borrowings, and in the number of borrowed nouns and noun phrases with plural referents. This situation leads to the conclusion that the unconventionality resulting from the omission of plural morphemes on English borrowings remains a peripheral phenomenon in present-day Romanian.

The factors which trigger the appearance of bare forms range from structural to social and psychological ones. Thus, the length and complexity of the transferred element, its phonological form and frequency of occurrence, as well as the bilingual ability and language attitudes of the borrowing language speakers can all be expected to influence the integration process. Although formal factors (constituent length and complexity) and frequency of occurrence have been shown to be relatively strong predictors of plural marking omission in bilingual discourse, we believe that socio- and psycho-linguistic factors should also be investigated in order to obtain a full account of bare forms in contact situations. Finally, in order to understand how English-origin bare forms challenge the Romanian linguistic norm it would be necessary to compare their distribution to that of native bare forms.

References

1. Avram, Mioara. 1997. *Anglicismele în limba română actuală*. Conferință prezentată la Academia Română. București: Editura Academiei Române.
2. Backus, Ad. 1996. *Two in One: Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: Tilburg University Press.
3. Baetens Beardsmore, Hugo. 1982. *Bilingualism: Basic Principles*. Clevedon, Avon, England: Tieto.
4. Boumans, Louis Patrick Clémens. 1998. *The Syntax of Codeswitching: Analysing Moroccan Arabic/Dutch Conversations*. Tilburg University Press.
5. Ciobanu, Georgeta. 2004. *Adaptation of the English Element in Romanian*. Timișoara: Editura Mirton.
6. Görlach, Manfred (ed.). 2005. *A Dictionary of European Anglicisms*. New York: Oxford University Press.
7. Guțu Romalo, Valeria (coord.). 2005. *Gramatica limbii române*. București: Editura Academiei Române.
8. Haugen, Einar. 1950. "The Analysis of Linguistic Borrowing". *Language* 26 (2): 211-231
9. Haugen, Einar. 1956. *Bilingualism in Americas: A Bibliography and Research Guide*. University of Alabama Press.

10. Jacobson, Rodolfo (ed.). 1990. *Codeswitching as a Worldwide Phenomenon*. Peter Lang.
11. Jacobson, Rodolfo (ed.). 2001. *Codeswitching Worldwide II*. Berlin, New York: Mouton de Gruyter.
12. Myers Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. New York: Oxford University Press.
13. Myers Scotton, Carol. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. New York: Oxford University Press.
14. Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin, New York: Walter de Gruyter.
15. Pandharipande, Rajeshwari. 1990. "Formal and functional constraints on code-mixing". In Jacobson, Rodolfo (ed.), 1990, pp. 15-31.
16. Poplack, Shana, David Sankoff, Christopher Miller. 1988. "The social correlates and linguistic processes of lexical borrowing and assimilation". *Linguistics* 26: 47-104.
17. Poplack, Shana, David Sankoff. 1984. "Borrowing: the synchrony of integration". *Linguistics* 22: 99-135
18. Stoichițoiu-Ichim, Adriana. 2001. *Vocabularul limbii române actuale: Dinamică, influențe, creativitate*. București: Editura All educational.
19. Treffers-Daller, Jeanine. 2001. "Contrastive sociolinguistics: Borrowed and codeswitched past principles in Romance-Germanic language contact". In Jacobson, Rodolfo (ed.), 2001, pp. 75-90.
20. Weinreich, Uriel. 1968. *Languages in Contact: Findings and Problems*. The Hague, Paris: Mouton.
21. *Colecția CD Capital 2005*. București: Ringier România.

Notes

ⁱ The annotated corpus was used as part of a PhD project on recent English borrowings in Romanian, conducted at "Babeș-Bolyai" University of Cluj Napoca

ⁱⁱ All these 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 Tufiș' supervision. The pos-tagged corpus is available at <http://corp.hum.sdu.dk/cqp.ro.html>