

LINGUISTICS AND COMPUTATIONAL LINGUISTICS

**MONITORING NEOLOGISMS IN NEWSPAPERS
WITHIN THE NEOROM PROJECT**

ANA-MARIA BARBU

*„Iorgu Iordan – Al. Rosetti” Institute of Linguistics, București – România
E-mail: anamaria.barbu@g.unibuc.ro*

This paper describes the contribution of the Romanian research team within the international project NEOROM. The project aims at building a lexical database consisting in neologisms which occur in newspapers written in all Romance languages since 2004. The paper has three parts. In the first one, we present the monitoring system adopted by the Romanian team, in the second one we describe the NEOROM interface, which allows to register the neological words in the database, whereas in the third part we give a short characterization of the words in the database, from the viewpoint of their formation.

Key words: neologisms, newspapers, Romanian, word formation.

1. INTRODUCTION

The NEOROM project (Neologisms in Romance languages) is an international undertaking designed to set up a multilingual database for Romance languages, consisting of recent words appeared in newspapers. Languages under monitoring are Catalan, French (spoken in France, Belgium and Quebec), Galician, European and Brazilian Portuguese, Romanian and Spanish. The process of inventoring new words started in 2004 and went on every year since then.

The project is under the care of the REALITER organization (<http://www.realiter.net>) and The Latin Union (<http://www.unilat.org>), and is conducted by Maria Teresa Cabré, professor at The Institute for Applied Linguistics of The University Pompeu Fabra, Barcelona, Spain. It mainly consists in a search engine for words inventoried in the databases corresponding to each language which is part of the project (Cabré 2004). The engine may be found at the address <http://obneo.iula.upf.edu/bneorom/index.php>.

NEOROM is one of the projects of the group L'Observatori de Neologia (<http://www.iula.upf.edu/obneo/obpresca.htm>), a group recognised as having the most prominent activity in the field of neologisms and terminology. Similar projects are

NEOSCOPE (<http://www.certa.usj.edu.lb/files/neoscope.htm>) and the *Oxford English Dictionary* division of the neologisms monitoring (<http://dictionary.oed.com/about/>).

According to the NEOROM terminology, neologisms under monitoring are lexical units formed out of one or several written words, which are not listed in a pre-established corpus of dictionaries, considered representative for the language in question¹. This is the so called 'exclusion corpus'. Words listed in the Romanian NEOROM database are not only those which, since 2004, appeared, for the first time, in newspapers, but also those which, for one reason or another, have been not included in dictionaries, although they were already in use.

On the other hand, the project pays equal attention to words which, though did not yet enter into dictionaries (and probably they never do) are still important, because of their capacity to witness about the contemporary productivity of the internal means of enriching the vocabulary, or, also, the individual lexical creativity. This essentially implies that what counts for the word selection is not the number *per se* of the occurrences in the newspaper corpus (even cases of hapax² are listed !).

Finally it ought to also be specified that the database contains word forms close to those occurring in dictionaries but which, though, underwent modifications not mentioned in dictionaries. It is about word forms with changed gender or changed part of speech, and also word forms with unattested plural or graphical shape.

The research reported here lies in the development of the Romanian database of NEOROM. In the next section we describe the monitoring system adopted to work out the list of words to be included in the database. In section 3 we give a description of the interface between the database and the slots which have to be filled for each listed word. Some of these slots represent search criteria within the base, so, their knowledge is necessary to all those interested in using the base. Section 4 is devoted to a short analysis of the results obtained for Romanian. The utility of the project and especially of the Romanian database is discussed in the section of conclusions.

¹ The term *neologism* does not have a unanimously accepted meaning, and for some linguists its use in this context may not seem adequate. Nevertheless, a discussion about the meaning of this term would exceed the limits of this paper.

² Hapax is a word with only one appearance in texts.

2. THE MONITORING SYSTEM

2.1 PRELIMINARY DATA

The corpus used to monitor the occurrences of recent words in newspapers consists in thousands of articles (about 14 millions words). They were collected from Internet, more precisely from the online version of several newspapers, such as *Adevărulonline*, *BBC-Romania*, *Cotidianul*, *Jurnalul Național*, *Libertatea*, *România Liberă* and others, appeared between 01.08.2004 and 15.10.2007. The corpus is accessible by courtesy of some Italian project-mates. It ought to be emphasized that the access to the online version of a newspaper, to the purpose of the subsequent automated processing of the text is not at all a simple task. This is why the assistance of our fellows was really valuable. Starting with this year we will also attempt to use the program *Buscaneo*, made available by our Spanish partners. Nevertheless we already have certain reservations, as far as the efficiency of this program is concerned. The main problem with this program is that, for each word not found by the program in the dictionary of the inflected forms, the program opens an entry in the database. Since our e-dictionary is relatively small, the number of not found words will be large, and so will be the number of the entries, too. Because of this large number of entries, we will not be able to apply automated or semiautomated procedures of selection to the words we dispose of so far.

As another procedure detail, it has to be mentioned that only *BBC-Romania* uses diacritics specific to written Romanian. The other newspapers do not. On the other hand, all the newspapers transcript the sound *î* as *â* (in a text without diacritics this means the use of the symbol *a*), except *Cotidianul*, which for the sound *î* uses the symbol *i* (in texts without diacritics).

Romanian dictionaries which form the exclusion corpus (that is, the reference work with respect to which a word is considered new³) are those e-dictionaries available at the address <http://dexonline.ro>. We quote, in this sense DEX '98, DLRA, DLRC, DLRM, DN, MDA and MDN, as well as DOOM2 because they have the advantage of the semiautomated search.

³ Recall that a word is considered new if it does not already occur in the exclusion corpus.

2.2 THE MONITORING PROCESS

The way in which neologisms are selected for the database is relatively simple, as it may be seen in Figure 1, below. At the same time, though, the process is laborious, due to the large number of data and, also, due to the diversity of the procedures to be applied (i. e. automated, semiautomated and manual ones). Articles in newspaper are indexed by a concordancer LUCON (<http://sourceforge.net/projects/lucon/>). Indexation yields inflected word forms which are further automatically searched in the e-dictionary RoMorphoDict of the inflected word forms (Barbu, 2008). This dictionary contains the complete paradigms of about 60000 lemmas in DOOM 1. Words that have not been found in the dictionary are stored in a separated list of the so-called ‘unknown’ words. An unknown form may be listed in it for one of the following reasons:

- the form in question does not lie in the reference e-dictionary (because this dictionary contains but 6000 lemmas, many of them not used in current speech);
- the form is an abbreviation, a measure unity or belongs to a class of arbitrary symbols;
- the form is a proper name;
- the form is a word from another language;

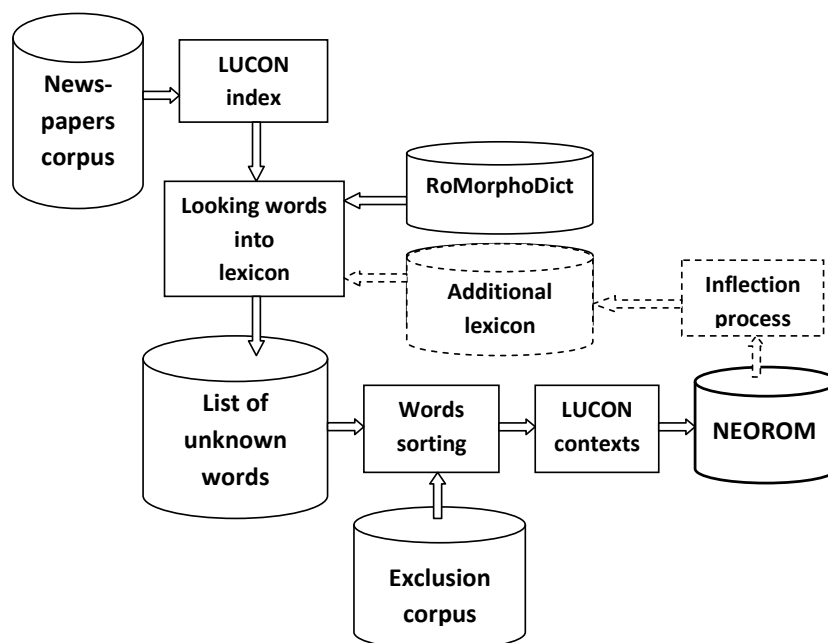


Fig. 1. Monitoring system

- the form is mistakenly uttered or written (as in the case of the word form *ostatec* with 111 occurrences);
- the form denotes temporal intervals or approximations (for instance *ianuarie-aprilie* “from January to April” *primăvară-vară* “spring-summer” *opt-nouă* “eight or nine”).

In the next step of the process, the list of the ‘unknown’ words is manually examined, in order to remove out of it what we consider to be the ‘residual words’. What remains after removing is confronted (also manually) with the above mentioned dictionaries, which constitute the exclusion corpus. The advantage with these dictionaries is that they may be found into an internet-accessible database, so that consulting them is a semiautomatic task. Each word not found in the exclusion corpus is associated with an identification card which allows the word in question to be introduced in the database NEOROM, as shown in the next section.

The process should be completed with a new step in which lemmas which have been already introduced into the NEOROM database receive inflected forms and are listed in a supplementary reference dictionary, in order to avoid to be inventoried one more time. Actually, the whole process might be improved if some conditions were met, namely if we disposed of a program able to download newspaper texts, if the exclusion corpus was in a proper electronic format, or if a program able to yield inflected forms of unknown words was available. Such programs should be integrated in a processing algorithm.

3. THE INTERFACE OF THE NEOROM DATABASE

In this section we describe the content of the card which has to be filled for each new word introduced into the base. As seen from Figure 1, the slots are the following.

General explanations about the word source:

Tipo: the proper information for this slot is that linguistic data are extracted from newspapers.

Lengua: one specifies that the language under investigation is Romanian

Ambito geografico requires the specification that the language subject to monitoring is Romanian spoken in Romania.

Fuente requires the specification of the document where the listed word is extracted from (for instance, *Libertatea* in its online form)

Publicacion indicates the date of the publishing (for example, January, 6, 2005)

The card also possesses a number of slots in which the listed word is analysed. The slots are the following.

Entrada gives the lemma.

Categoria gramatical supplies information such as part of speech, gender (for nouns or adjectives), the verb type and so on. All these pieces of information have to be chosen from a pre-established list. We initially coped with a problem concerning the gender: the neuter gender did not appear among the values of the gender category, because in the other Romance languages the neuter does not exist any more. The problem was subsequently solved.

Contexto: Each lemma has to have mentioned a context. It is the context in which the word occurs in the newspaper. In most cases, a word occurs in several newspapers. In this situation, one chooses the earliest occurrence. Nevertheless, one may also make the decision to choose the context which gives the clearest idea about the meaning of the word (one may find, for instance, contexts where the meaning of the word is explained). Contexts are extracted with the aid of the concordancer LUCON. Among several facilities provided by LUCON, there is also the advantage of supplying in a simple manner the source of the context (that is, the title of the newspaper and the date of its publishing). These pieces of information must appear in the card.

Aspectos tipográficos: This heading refers to the manner in which a word is highlighted. Usually, highlighting is done through commas or brackets. Nevertheless there are other ways in print texts, as well, which are not available in our electronic versions.

Tipo de neologismo: This slot is designed to specify the word formation. There are several types of word formation, given in a pre-established list with corresponding identifiers. The content of this list, as well as a detailed analysis of the database inventory are presented in (Barbu 2010).

Sección del periódico: In this slot one indicates the field of the article from which the word was extracted. Typical fields of a newspaper are politics, economy, finances, society, commerce, culture, health, sports, entertainment. Usually, in the online version

of a newspaper, the field of an article is not specified. This is why the field is deduced from the content of the article.

PLATAFORMA NEOROM Francès · Galleg · Portugués

Bienvenido/a *ivr* | Gestión de fichas :: Prensa - Editar ficha

- Gestión de fichas
- Edición de fichas
- Consulta realizadas
- Consulta general
- Consulta impresión
- Salir

Datos generales del vaciado

Tipo: Prensa
 Lengua: Rumano
 Ámbito geográfico: ROMANIA
 Fuente: Libertatea
 F. Publicación: 06/01/2005 (dd/mm/aaaa)

Datos de la ficha

Entrada:

Categoría gramatical: Nombre femenino

Contexto: Valentina Pelinel: La o prezentare, *designera* nu mi-a inchis bine gaicile la fusta si am ramas fara ea.

Aspectos tipográficos: Sin marca tipográfica

Tipo de neologismo: SINT - Neologismo sintáctico

Sección del periódico: CULT - culturale

Página:

Autor: Sí No

Código nota: 2g


Nota: variante sintáctica de s.m. „designer”

Estado: Incompleta Completa

Comentarios:

Historial de cambios

	Creación	Completado	Corrección	Validación
Autor	ivr	-	-	-
Fecha	20/01/2009	-	-	-



NEOROM
RED DE OBSERVATORIOS DE
 NEOLOGISMOS DE
 LENGUAJA DE LAS COMUNIDADES EUROPEAS

Observatori de Neologia · Institut Universitari de Lingüística Aplicada · Universitat Pompeu Fabra
 Roc Boronat, 138, 08018 Barcelona. Tel. 935 422 322 Fax 935 422 321
 observatori.neologia@upf.edu

Fig. 2. Entry card in database NEOROM

Página: Informations about the page are available only if the newspaper is in its print form. The online version does not supply the page of the article.

Autor confirms if the name of the author of the article is known, or alternatively, the name of the person quoted in the article as having been used the word listed in the card.

Código nota: The card also contains a pre-established list of additional explanatory specifications concerning neologisms. If necessary, the neologism under analysis may receive a supplementary characterization, which represents the content of the rubric *codigo nota*.

Nota represents the proper explanatory note of the neologism. The noun *designeră*⁴, in Figure 1, for instance, is a syntactic neologism (for the definition of a syntactic neologism, see the next section). It is formed by the change of the gender of the noun *designer* from masculine to feminine.

Estado: The extent to which the card is filled is evaluated by the pair Complete/Not complete, terms which are designed to occupy the slot *estado*.

Comentarios: The operator who fills the slots is allowed to make comments about the card properties. Comments are not made public.

Historial de cambios: The system automatically specifies the name of the operator who worked, modified, or validated the card, as well as the date of the operation.

The process of filling the card ends when the option *Guardar* is activated. One may get the same result by leaving the card unchanged. This is done by pushing the button *Cancelar*. Card deletion from the database is done with the option *Borrar*.

4. SHORT ANALYSIS OF THE RESULTS

So far the NEOROM database contains about 2000 words. We intend that till the end of 2010 the inventory increases to more than 4000 lexical units. At the moment, one can enrich the database with about 100 entries a day. We had to classify the words according to their formation. Modes of word formation are standardized and common to all languages under investigation. In order to observe this standard, we had to make some compromises, which will be described below.

- Word classification according to the mode of formation relies on the analysable form of the words, not on their etymology (for instance, some analysable words may actually be loan words).

⁴ “designer”, in the feminine

▪ If several modes of formation may be invoked to explain a certain word formation, we opted for only one. In principle, we took into account the words found in exclusion corpus as derivation bases. For example, the neologism *coprezentatoare*⁵ was considered a word derived by prefixation from the base word *prezentatoare* found in exclusion corpus, rather than a suffixal derivation from the base *coprezentator*⁶, because *coprezentator* is not in reference dictionaries. Another example: the word *sepepist*⁷ is formed by both suffixation and abbreviation. Nevertheless, only the former mode of formation has been retained into the database.

▪ Some modes of word formation differ from what reference works for Romanian propose. For instance, in NEOROM, the feminine word *designeră* is characterized as a syntactic (or grammatical) neologism, whereas in the Romanian works the word is considered as being obtained by suffixation. Another example of the same type is supplied by the word *superblatist*⁸. In NEOROM, this word is taken as being formed through both suffixation and prefixation. Nevertheless, Romanian works consider that this mode of formation characterizes a quite different class of derived words, those which usually do not have a full word base but merely a root base (for example, *a îmbrăca*⁹).

▪ The NEOROM framework considers that apparent prefixes, which in their languages of origin were adverbs or prepositions (*anti-*, *extra-*, *intra-*) are true prefixes, whereas the others (ex. *auto-*, *agro-*, *bio-*) create compounds (that is, multi-word expressions).

▪ Unfortunately, the NEOROM framework does not incorporate two further modes of word formation, namely backward derivation and derivation by substitution of affixes. Consequently, cases like *atotputere* < *atotputernic*¹⁰ (affix substitution) and *bâlbă* < *bâlbâi*¹¹ (backward derivation) cannot be adequately analyzed.

⁵ “co-announcer”, in the feminine

⁶ “co-announcer”, in the masculine

⁷ member of the service for the protection of the dignitaries

⁸ a slang word meaning a person who uses public means of transportation without paying the ticket

⁹ “to get dressed”

¹⁰ might<mighty

¹¹ stammering<to stammer

Apart from NEOROM canonical indications concerning word formation, the main reference works for this topic in Romanian employed in our research were (Coteanu et al. 1985) and (Stoichițoiu-Ichim 2001). Nevertheless, it ought to be said that the characterization of the formation of the neologisms listed into the database does not always represent the conclusion of a rigorous and prolonged research. This leaves room for subsequent corrections.

In what follows, we shortly present a partition of the neologisms from the Romanian database from the viewpoint of the most representative modes of formation. We use symbols taken from the platform NEOROM.

Derivation with affixes (27,5%)

Words derived with suffixes (FSUF: 13,54%)

The most frequent suffixes are **-ist** (ex. *cablist*¹², *galerist*¹³, *forumist*¹⁴), **-re** (*fumare*¹⁵ *gudurare*¹⁶, *microcipare*¹⁷), and **-(e/i)an** (ex. *becalian*,¹⁸ *cațavencian*¹⁹). An unusual suffix is **-ing** (*bancheting*²⁰, *cuponing*²¹).

Words derived with prefixes (FPRE: 13,96%)

No doubt, the most productive prefix is **ne-**. Words derived with it represent 30% from the total of the words that are prefix-derived (for example, *nealegere*²², *necalitativ*²³, *nesustenabil*²⁴); **ne-** is followed by the prefix **anti-** (22,5%). It derives adjectives from nouns (*anticutremur*²⁵, *antisărăcie*²⁶, *antisforăit*²⁷). Another prefix preferred by speakers is **cvasi-** (*cvasi-absență*²⁸, *cvasiinațional*²⁹, *cvasi-război*³⁰).

¹² TV cable company worker

¹³ owner of an art gallery

¹⁴ user of the internet forums

¹⁵ “smoking”

¹⁶ “fawning”

¹⁷ the action of implanting a microcip

¹⁸ from the proper name “Becali”

¹⁹ from the proper name “Cațavencu”

²⁰ the organization of the symposia

²¹ a tool of marketing, which uses vouchers

²² the fact of not being elected

²³ low quality (or even no quality)

²⁴ unsustainable

²⁵ anti-earthquake

²⁶ anti-poverty

²⁷ anti-snoring

²⁸ literally, “quasi-absence”

Compounds (27,9%)

One has to firstly distinguish words compounded out of full words (FCOM: 17%). This mode of composition is used to currently express political, economical, organizational or social relationships, such as *americano-canadian*³¹, *chino-ruso-indian*³², *etnico-religios*³³ *administrativ-teritorial*³⁴, *cerere-ofertă*³⁵, *financiar-contabil*³⁶. It is also employed to denote hierarchical positions or occupations: *director-coordonator*³⁷, *cazangiu-tinichigiu*³⁸ etc.

Elevated compounds (FCULT: 8,8%.)

This type of compounds is obtained by means of apparent prefixes (others than former prepositions or adverbs). The most frequent apparent prefix is the neologism **euro-** (*euroasistat*³⁹, *eurobarometru*⁴⁰, *eurogară*⁴¹, *europubelă*⁴²). Another well-known apparent prefix is the older **auto-**, having the meaning “by itself”, “from itself”: (ex. *autoadresa*, *autobronzant*, *autoironic*). Very recent, too, are **tech-** (*tech-cultură*, *tech demo*) and **web-** (*web designer/web-designer*, *web hosting*, *website*), both from English.

Compounds through abbreviation (2,1%) gave rise to three types of words: abbreviations according to initials, (shortly, initials, FTSIG), acronyms (FTACR) and truncations (FTABR). Pure initials are very rare: *sereleu*⁴³ or *sms*⁴⁴. What is dominant instead is the combination between abbreviation and suffixation. In this respect, the most used suffix is -ist: *cederist*⁴⁵, *cekist*⁴⁶, *sepepist/sepepistă*.

²⁹ literally, “quasi-national”

³⁰ literally, “quasi-war”

³¹ “American-Canadian”

³² “Chinese-Russian-Indian”

³³ “ethnic-religious”

³⁴ “administrative-territorial”

³⁵ “request-offer”

³⁶ financial account

³⁷ executive director

³⁸ coppersmith and tinsman

³⁹ person socially assisted according to the assistance rules of the European Union

⁴⁰ “eurobarometer”

⁴¹ “euro-railway station”

⁴² literally, “euro-recycling bin”

⁴³ From SRL

⁴⁴ abbreviation for “short written message”

⁴⁵ member or sympathizer of the CDR (The Democratic Convention from Romania)

⁴⁶ member of the CK (the ancestor of KGB)

On the other hand, truncation usually applies to elevated compounds which contain apparent prefixes. It is these apparent prefixes that are kept in the truncated words, which are used as invariable adjectives and, subsequently, as nouns: *combo, eco, ego, electro, etno, hidro, homo, lesbi, macro, mix, termo, turbo, vice*.

Conversion (FCONV: 4,4%)

Conversion yields a new word as a consequence of the fact that the word changes its original part of speech, without changing its base form. The most typical case of conversion yields adjectives from nouns. Some semantic conditions of this type of conversion are the following.

(i) A person may be identified due to one of his/her characteristics: *anchetat*⁴⁷, *mascat*⁴⁸, *național*, *necăjit*⁴⁹, *penal*⁵⁰, *vestic*⁵¹ etc.

(ii) People (usually, groups) may be identified by means of the symbolic colour(s) of their collectivity. This especially happens with political or sports collectivities (*verzii, alb-albaștri, alb-violet, roșu-albaștri, tricolori, vișinii*⁵² etc).

(iii) Abstract properties (such as to be beautiful, essential, divine, economical) come to denote entities, when the adjective is articulated in the singular: *derizoriul, divinul, economicul, eroicul, eroticul, esențialul, eticul*⁵³.

(iv) Conversion also takes place, when in a phrase of type noun-adjective the noun is suppressed. Under these conditions, the adjective becomes a noun: (*alegeri*) *anticipate, generale, legislative, locale; (aparate) electronice, (medicamente) compensate; (scor) egal*⁵⁴ etc.

Syntactic (or: grammatical) neologisms (SINT: 2.3%)

Syntactic neologisms are new words which exhibit a change in one of their characteristic grammatical categories. The following situations are documented.

⁴⁷ person under criminal investigation

⁴⁸ fighter in the special law enforcement, who wears a masque, in order to protect his identity

⁴⁹ a person with modest social standing

⁵⁰ person suspected of violating the rules of the criminal law

⁵¹ person coming from a Western country

⁵² the greens (for the ecologists), the red-blue (for the football players of the team Steaua) etc.

⁵³ literally, derisory>the derisory, divine> the divine, heroic> the heroic etc.

⁵⁴ anticipated (elections), subsidized (medicine) etc.

(i) Masculine nouns earn a pair which is feminine: *administratoră, arbitră, doctoră, fană, stelistă*⁵⁵.

(ii) Neuter nouns in the plural earn a masculine plural pair: *amortizori, boscheți, flotori*⁵⁶

(iii) Nouns without plural earn a plural form : *bejuri, designuri, disconforturi, euroi*⁵⁷

Graphical and morphological doublets (FVAR: 9,3%)

What we consider a doublet within the NEOROM framework is especially words spelled out differently from what the literary norm requires: *aiatolah* (the literary form, *ayatolah*), *afrodisiac* (*afrodiziac*), *chic* (*șic*), *click* (*clic*), *boutique* (*butic*). Another case of graphical doublet are words formed with apparent or true prefixes, which are spelled out with hyphen (the norm recommends the spelling out without hyphen).

As for morphological doublets, they are words with number or case forms, different from the literary ones: *aeropoarte* (the literary form, *aeroporturi*⁵⁸), *cenușei* (*cenușii*⁵⁹), *gazoane* (*gazonuri*⁶⁰), *girofare* (*girofaruri*⁶¹).

English loan words (ME: 19%)

A fifth from the total number of the neologisms in the period under scrutiny comes from English. There are some terminological domains in which the presence of English loan words is really impressive.

Professions and occupations: *advertiser, account manager, chief executive officer, chief marketing officer, corporate director, developer, entertainer, fashion coordonator, gambler, graffer, head-hunter manager, sound-designer, web designer*.

Sports: *cross country, down hill, four-cross, full-contact, greencard, goal-keeper, hacky*.

⁵⁵ manager (masculine)>manager (feminine), referee (masculine) > referee (feminine), doctor (masculine) > doctor (feminine) etc.

⁵⁶ shock absorbers (neuter, plural) > shock absorbers (masculine, plural), bushes (neuter, plural)> bushes (masculine, plural) etc.

⁵⁷ design>designs, discomfort>discomforts, euro>euros

⁵⁸ “airports”

⁵⁹ “ash” in the genitive/dative

⁶⁰ “turf” in the plural

⁶¹ identification traffic lights used by the cars of the police, medical emergency and fire brigades.

Music: *backing vocal, boy-band, cover, dance* (adj.), *(deep) house, death-metal, downtempo, drum & bass, electric-gipsy, featuring, funk, heavy rotation*.

Fashion: *baby doll, face-painting, fashion, fusion-fashion, gloss, hair-styling*.

Electronics/computers: *dial-up, downloada, enter, gateway, hands-free, hi-tech*.

Food: *fresh, junk-food, light*.

There are, of course, loan words from other languages, as well, but their number is not significant. One may also document the existence of adapted loan words (that is, words that do not preserve any more their original orthography), yet they, too, are little numerous. Their small number may be explained by the preference of some speakers for the original orthography, despite the fact that the adopted orthography has been already normed (see, for example the section of the orthographical doublets).

Idiosyncratic loan words (Xenism: 0,5%)

By 'idiosyncratic loan words' (abbreviated, ILW), we term loan words which refer to phenomena and facts not specific to the Romanian culture. Examples of ILWs are *datsan* („trend in budhist religion”), *fatwa* („law originating in the sacred texts of the muslims”), *kumara* („technique of agricultural labour in the Easter Island”), *guerilleros* („guerilla fighters”), *mzungus* („man of white race”).

Others (A: 5%)

There are words the etymology of which is not clearly determined. These are words coming from dialects, slang or professional jargons, which, though, appeared in mass media. Some words from the medical and pharmaceutical terminology belong to this category (*aflatoxină, botox, carmol*), but also words without obvious etymology (*amărăștean, amoform, aszu, critifiction, flex a. o.*).

5. CONCLUSIONS

It has been apparent from the above considerations that the present study touched several aspects of the issue of the neologisms. It represents the first systematic monitoring of the Romanian written press, thanks to the electronic resources. For this reason, this research could represent a touch stone for a subsequent monitoring. We are referring to the fact that the aim of the NEOROM project is not the achievement of an

inventory of the neologisms in mass media, but the achievement of an inventory of the words used in newspapers in a certain year. Or such an inventory could not be obtained in a sufficiently exact manner without drawing a demarcation line at a certain moment, and say “here is the inventory of the words existing in the language, so far”.

In the second place, this paper is relevant from a lexicographical point of view. In this respect, it was noticed that a great number of words have been not yet registered into dictionaries, despite their frequency in use or their ‘long’ presence in the language. Most of them are members of words families, which are only in part registered into dictionaries (for example, one may find in dictionaries the word *minimalist*, but not also *minimalism*. From now on, this drawback can be eliminated.

At the same time, there are words in our database, which are also present in more or less specialized dictionaries (of slang, economy, computer sciences, sports, as well as dictionaries of synonyms, antonyms, or ortographical dictionaries etc). The fact that these words do appear in mass media show that they are, to a certain extent, accessible to the majority of the speakers; and for this reason, they would deserve to be listed in general dictionaries, too, especially if they exceed a certain frequency threshold. In this sense, it ought to be said that, surprisingly, words of wide circulation (like *denim*⁶² or *sclipici*⁶³) are not registered in the general dictionaries. But even if not each word from our database should be introduced into a dictionary of Romanian, the present study adduces new evidence concerning contemporary lexical creativity and may be helpful in prospecting evolution trends of the contemporary Romanian. Also, the research paves the way for a comparison between the lexical dynamics of Romanian and the one of the other Romance languages.

Acknowledgements. I wish to thank Ioana Vintilă-Rădulescu, the coordinator of the NEOROM Romanian team, for her constant assistance in my effort of improving the quality of the research. I also thank the anonymous reviewers of this paper, for their comments.

⁶² name obtained by antonomasia for the products of the DENIM company

⁶³ shimestone

REFERENCES

1. (Barbu, 2008) BARBU A.M., *Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries*, Proceedings of Language Resources and Evaluation Conference – LREC 2008, Marrakech; <http://www.lrec-conf.org/proceedings/lrec2008/>, 2008.
 2. (Barbu, 2010) BARBU A.M., *Baza de date românească din cadrul platformei Neorom*, *Limba română*, 2010, 2, București (in print).
 3. (Cabr , 2004) CABR  M. T. (coord.), *Metodolog a del trabajo en neolog a: criterios, materiales y procesos*, Observatori de Neologia, Papers de l'IUULA, S rie Monografies, 9, Universitat Pompeu Fabra, Institut Universitari de Ling istica Aplicada, Barcelona, 2004.
 4. (Coteanu et al., 1985) COTEANU I., FORĂSCU N., BIDU-VRĂNCEANU A., *Limba română contemporană. Vocabularul*, Editura Didactică și Pedagogică, București, 1985.
 5. (Stoichițoiu-Ichim, 2001) STOICHIȚOIU-ICHIM A., *Vocabularul limbii române actuale. Dinamică, influențe, creativitate*, Editura All, București, 2001.
- DEX '98 = *Dicționarul explicativ al limbii române*, Editura Univers Enciclopedic, București, 1998.
- DLRA = Zorela Creța, Lucreția Mareș, Zizi Ștefănescu-Goangă, Flora Șuteu, Valeriu Șuteu, *Dicționar al limbii române actuale* (ediția a II-a revăzută și adăugită), Editura Curtea Veche, București, 1998.
- DLRC = Vasile Breban, *Dicționarul limbii române contemporane*, Editura Științifică și Enciclopedică, București, 1980.
- DLRM = *Dicționarul limbii române moderne*, Editura Academiei, București, 1958.
- DN = Florin Marcu și Constant Maneca, *Dicționar de neologisme*, Editura Academiei, București, 1986.
- DOOM 1 = *Dicționar ortografic, ortoepic și morfologic al limbii române*, Editura Academiei Republicii Socialiste România, București, 1989.
- DOOM 2 = *Dicționar ortografic, ortoepic și morfologic al limbii române*, ediția a II-a, Editura Univers Enciclopedic, București, 2005.
- MDA = *Micul dicționar academic*, Editura Univers Enciclopedic, București, 2002.

MDN = Florin Marcu, *Marele dicționar de neologisme*, Editura Saeculum, București, 2000.

Received August 16, 2010