

Initial combining forms across registers: the case of *aero-*, *hydro-* and *cyber-*.

Begoña Montero-Fleta
Carmen Pérez-Sabater
Universitat Politècnica de València
Department of Applied Linguistics

Abstract

This study investigates to what extent certain lexical differences can be attributed to factors such as register. To this end, the paper addresses the empirical domain of a word-creation process by means of initial combining forms (ICFs) of neoclassical origin. The research undertaken is based on the use of ICFs in the *Corpus of Contemporary American English*, a corpus compiled from texts drawn from spoken, fiction, magazine, newspaper and academic sources. The study focuses on the variation of neoclassical elements across registers, and questions their morphological status and currency. The findings confirm the hypotheses of the higher presence of initial combining forms in science and technology domains in magazine and academic journals, and the lower use in spoken registers. It also confirms that nowadays some initial combining forms have developed into free morphemes, a fact which contradicts the essence of the bound nature of combining forms. The study highlights that neoclassical forms are a salient resource in the lexical expansion of terminology and claims their pedagogical implications in ESP word-formation processes.

Keywords: register variation, initial combining forms, neoclassical morphemes, morphology, terminology.

1. Introduction and justification of the study

This study addresses the presence of complex words with a neoclassical component across different registers. Registers have been scholarly classified in terms of particular contexts of language use (Ferguson 1982; Zwicky and Zwicky 1982; Biber 1995; 2012). Accordingly, to undertake a comprehensive linguistic analysis of registers, consideration of a representative selection of linguistic features is required to see the differences in the type of language. In this context, one of the core lexical characteristics proposed by Biber (1995) that serves to distinguish differences across genres is lexical specificity; this is sometimes related to the need that languages frequently experience to update the existing stock of words so as to cater to new concepts in different specialized areas (Author 2011).

One of the processes of creating new words is by combining elements: two or more constituents are used to make these elements work as one word to name new concepts. The present study focuses on initial combining forms, a word-formation resource as a result of a conscious, creative process, in particular for scientific terminology (Amiot and Dal, 2007). A controversial issue regarding initial combining forms (hereafter ICFs) is the different names they receive. ICFs are stems from neoclassical origin which have lexical meaning and form compounds. They are also called pseudo-prefixes (Bauer 1983, Lüdeling et al. 2002; Kastovsky 2009) or affix-like initial combining forms

(Bauer 1983). Their etymology is voiced in other names such as neoclassical prefixes, classical compounds, classic bound morphemes, classic prefixes or neoclassical words (Lüdeling et al. 2002). Other names such as confixes, or formatives (Lüdeling 2006), bound stems (Selkirk 1982; Scalise 1986 or Ralli 1988, 2007), affixes (Williams 1981), or semi-words (Martinet 1979) are also common. Adams (2001) refers to ICFs as quasi-lexemes. Lexicographic sources adopt other criteria when naming ICFs in dictionary entries; a wide naming variation is observed across dictionaries, from combining forms to prefixes, or simply word-elements. Considering the diversity of terminology associated with ICFs, the present study adopts the term initial combining forms after the designation of the *Oxford English Dictionary* in its third edition: “a combining form differs from a prefix or suffix by having a relatively full lexical meaning” (see McCauley 2006), and also used by Bauer (1983), Lüdeling et al. (2002), Prčić (2005) and Kastovsky (2009).

It is striking to observe that Marchand (1969) in his ground-breaking treatise on English word-formation, *The Categories and Types of Present Day English Word-Formation*, did not study ICFs, neither did Alvar (1993) in his seminal research on Spanish word-formation. Some scholars, as Bauer (1998), Kastovsky (2009) and Lüdeling (2006), among others, claim that ICFs have never been researched systematically. However, the topic is recently starting to draw scholarly attention: Prčić, T. (2008) draws a systematic dividing line between prefixes and initial combining forms in English. On the other hand, Díaz-Negrillo (2014) analyses neoclassical compounds and combining forms. Scher & Nóbrega (2015) discuss the structural parallelism present in neoclassical and stem-based compounds also in English. But the morphological process under study is not exclusive of English; combining forms are used today in most European languages. Petropoulou (2009a, 2009b) compares combining forms in English and Greek; Ralli (1988, 1992, 2008) studies ICFs in Modern Greek; Corbin (2001) addresses ICFs in French; Iacobini (2004) and Montermini (2002) in Italian, Lüdeling et al. (2002) in German, and San Evaristo (2010) focuses on Spanish. These studies show that not all European languages have been equally receptive to classical combining forms. Some languages have used them in combination with other classical elements, while others have been more prone to combining them with native elements, specifically when creating their own technical vocabulary, as is the case of German.

The point of departure of the present study is Baayen’s (1994) and Evert and Lüdeling’s (2001) postulate that word-formation patterns are highly dependent on text type, and that “neoclassical words, for example, are much more likely in scientific texts than in everyday language” (Evert and Lüdeling 2001: 17). The research undertaken forms part of a more extensive ongoing project on the plethora of combining forms. An earlier publication of the author on word-creation patterns using traditional and prepositional prefixes (Author, 2013) showed the lack of consistency among scholars in the treatment of prefixation processes, mainly when dealing with word-formation elements which appear to blend into each other.

The present manuscript is structured as follows. Section 2 introduces the aims of the study and states the research questions and hypotheses. Next, Section 3 addresses the methodology followed. To answer the research questions posed, Section 4 analyses the ICFs chosen for the study separately, and highlights the data obtained on their frequency and use. The appraisals of the findings are then discussed in Section 5, followed by the conclusions and pedagogical implications in Section 6.

1. Aims of the study

The study aims at answering the following research questions:

Research question 1: Do ICFs show differences in their use across registers?

Hypothesis 1: As ICFs are learned words, they are associated with science and technology debates and, thus, appear more frequently in academic journals and specialized magazines.

Research question 2: Are ICFs bound or free constituents?

Hypothesis 2: The assumption of the present study is the bound nature of ICFs, a feature postulated by previous approaches that claim that ICFs cannot stand alone as free words (Bauer 1983, Fischer 1998, among others).

Research question 3: Are ICFs subject to frequency variation across registers?

Hypothesis 3: As many ICFs name scientific or technical concepts of current use, no variations should be expected.

3. Methodology

Following Biber (1995), analyses of register features should be necessarily quantitative to see the differences from the relative distribution of linguistic features. A corpus-based approach to the study of register variation may reveal regularities in language use that are otherwise difficult to find. Thus, to analyze combining forms in current language in use, the *Corpus of Contemporary American English* (hereafter *COCA*), developed by Mark Davies from Brigham Young University (2009), was employed. This corpus has been collected to study language variation across registers. *COCA* is composed of more than 450 million words from more than 160,000 texts; 20 million words more each year are included from 1990 to 2012. The corpus is evenly divided into the following five domains: oral, fiction, magazines, newspapers and academic journals¹. The present analysis was based on the entries of a random sample of three ICFs, *aero-*, *hydro-* and *cyber-*, in the different registers, each register containing approximately 81 million words. *COCA* provided the frequency of each combining form sought, as well as concordance and collocation facilities to study words in context.

3

4. Data analysis and findings

In this section, the ICFs chosen for the study are dealt with separately to facilitate the response to the three research questions posed and to highlight the data obtained on their frequency and use.

¹ See *COCA*'s webpage: <http://corpus.byu.edu/coca/>

4.1. Aero-

The *Online Etymology Dictionary* defines *aero-* as “comb. form meaning ‘air, atmosphere’, from Gk. *aero-*, from *aer* (gen. *aeros*) air, lower atmosphere”.

4.1.1. The following table shows the token frequency of *aero-* in *COCA*:

Table 3. *Aero-*: Total entries and rate per million in different registers (*COCA*).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	11,001	707	894	4,162	2,124	3,114
Per million	23.71	7.40	9.89	43.55	23.16	34.19

A higher use of *aero-* was documented in magazine sources (37.83%), followed by academic registers (28.30%), and newspapers (19.30%). Spoken registers attested significantly lower figures (6.42%). A frequency-ordered distribution search revealed the types with the highest token frequency²: *aerospace* (2,130), *aerobics* (1,199), *aerodynamic* (635), *aeronautics* (506), *aero* (483), *aerosol* (485), *aerodynamics* (379), *aeronautical* (298), *aerosmith* (236), *aeros* (128), *aeroflot* (123), *aeronautic* (99), etc. A high number of nouns was observed in *aero-*combinations, but adjectives, such as *aerodynamic* or *aerostatic*, were also involved.

The terminological unit yielding the highest token frequency in *COCA*, *aerospace*, was then analysed and the following results were documented:

Table 4. *Aerospace*: entries and rate per million in different registers (*COCA*).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	2,130	233	115	645	625	512
Per million	4.59	2.44	1.27	6.75	6.81	5.62

As Table 4 illustrates, magazines recorded the highest number of *aerospace* entries (30.28%), these values were closely followed by newspapers (29.34%), and academic sources (24.03%). Spoken registers reached significantly lower (10.93%). Fiction texts ranked the lowest rates (5.39%). Examples (1) show this neoclassical compound in context:

- a. *Northrop Grumman, a global aerospace and defense technology company, recently announced plans to close troubled shipyards and leave the shipbuilding industry (MAG. Saturday Evening Post, COCA 5, 2012)*³
- b. *we'll keep a close eye on private space companies as they continue to make aerospace history (MAG. Mechanics, COCA 3, 2012)*

² Data are shown in decreasing order with indication of the number of tokens found.

³ The examples from *COCA* throughout the paper are followed by the information in the corpus referring to the source (e.g., MAG) and the topic being discussed (e.g., bicycling). The number after *COCA* corresponds to the identification of the entry in the corpus to which the example relates; the year of publication is indicated next.

Aerospace, as most *aero*-compounds in the corpus under study, was mainly used as a solid compound. Alternative spellings of *aerospace* were not representative because of their minimal occurrence: *aero-space* was recorded only five times, and *aero space* was used just twice.

4.1.2. Most studies have highlighted the bound state of ICFs, a characteristic which is reflected in the definition of ICFs in *The Oxford English Dictionary*. Relatedly, some studies have claimed that with the addition of an appropriate inflectional ending, ICFs may become free units (Ralli 1992; 2008). However, the analysis of the global tokens of *aero*- shown in Table 4 above revealed the use of *aero* as a stand-alone element. Table 5 displays the number of occurrences found:

Table 5. *Aero*: entries as a stand-alone token and rate per million (COCA).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	483	10	21	345	52	55
Per million	1.04	0.10	0.23	3.61	0.57	0.60

Aero was documented as a recurrent, autonomous element in current registers with an outstanding imbalance in favor of magazine sources (71.42%); no addition of inflectional ending elements was attached to it. Entries for the free morpheme *aero* in magazines were mainly recorded from specialized contexts in magazines describing technical issues about wind resistance as a result of aerodynamically optimized designs or more favorable aerodynamic positions.

- (2)
- a. *Don't think about getting low and **aero** on this bike; that's not what it's made for* (MAG. Bicycling, COCA 67, 2011)
 - b. *Tolpadi is currently the technical leader of **aero** and acoustics in generator engineering at GE-Energy* (ACAD. Mechanical Engineering, COCA 93, 2011)
 - c. *Everyone wants **aero**, but at this price range, deep metal rims add weight for negligible* (MAG. Bicycling, COCA 41, 2012)
 - d. *The **aero** comes with a 6-speed manual, the arc a 5-speed* (MAG. Bicycling. COCA 251, 2003).

However, it is important to point out that a manual analysis of the stand-alone entries in the corpus questioned some apparently free *aero* morphemes. Some of them were later discarded as free morphemes, as they showed to be the first constituents of compounds which formed polylexical terminological units despite their spaced spelling. The bound status of the compounds was evidenced in examples as the following:

- (3)
- a. *Swap in a 78-degree post for a position that is optimized for clip-on **aero** bars (the stock FSA SL-K carbon handlebar is reinforced for this purpose)* (MAG. Bicycling, COCA 5, 2012)
 - b. *Caveats aside, wider **aero** wheels are here to stay and could come to dominate the market* (MAG. Bicycling, COCA 25, 2012)

4.1.3. Further research on the use of *aero-* over time revealed a lower use of this ICF in the last decades; 2,124 occurrences were documented from the period 1990-1994 vs. 1205 from 2010-2012, thus exhibiting a current 43% lower use.

4.2. *Hydro-*

The *Online Etymology Dictionary* provides the following definition for *hydro-*: “*Hydro-*: before vowels *hydr-*, word-forming element meaning ‘water’, from Gk. *hydro-*, comb. form of *hydor* ‘water’”.

4.2.1. The global number of occurrences of *hydro-* revealed the following figures:

Table 6. *Hydro-*: Total entries and rate per million in different registers (COCA).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	13,280	987	939	4,683	1,531	5,140
Per million	28.62	10.33	10.38	49.01	16.69	56.44

Academic sources documented 38.70% of *aero* occurrences, while magazines amounted to 33.93%. Newspapers stayed far behind (11.52%); spoken and fiction registers obtained a scarce number of results (7.43% and 7.07% respectively).

A frequency search for *hydro-* combinations recorded the terms with higher number of entries: *hydrogen* (5470), *hydrocarbons* (692), *hydroelectric* (649), *hydrocarbon* (410), *hydropower* (355), *hydro* (355), *hydrology* (249), *hydrological* (246), *hydrologic* (212), *hydrothermal* (200), *hydrogenated* (195) or *hydroponics* (150). Although the most frequent *hydro* forms were nouns and adjectives, also adverbs were attested, as *hydroponically* (31). *Hydrogen*, the term with the highest number of entries containing *hydro* in COCA, was then analyzed to compare its presence in different registers. The following number of entries was attested in all five registers:

Table 7. *Hydrogen*: entries and rate per million in different registers (COCA).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	5,470	584	433	2,181	538	1,736
Per million	11.79	6.11	4.79	22.82	5.87	19.06

Hydrogen ranked slightly higher in magazines (39.85%) than in academic sources (31.72%). Spoken registers obtained lower entries (10.76%). An analysis of the tokens of *hydrogen* in magazine and academic sources demonstrated their intrinsic science and technology character. Some examples illustrate the use of *hydrogen* in context:

(4)

- a. Stars formed of elements other than **hydrogen** would have different masses and evolutionary histories than normal hydrogen-rich stars (MAG. Astronomy, COCA, 26, 2012)
- b. Indoor air pollutants sulfur dioxide (SO₂), nitrogen dioxide (NO₂), **hydrogen** sulfide (H₂S), formaldehyde (HCHO), carbon monoxide (CO), and particulate matter were measured using passive samplers over a 7-day period (ACAD. Environmental Health Perspectives, COCA, 54, 2012).

It is worth mentioning that orthographic alternation was observed between the hyphenated term *hydro-gen*, and the term with the spaced format *hydro gen*, both with only 1 occurrence vs. its use as a solid compound, e.g. a compound whose components are solid in printing or writing: *hydrogen*, which showed 5470 tokens. The high use of the compound has resulted in an entrenchment of the word in the lexical memory of the community and, thus, its spelling as a solid compound. However, other compounds with a comparatively lower number of tokens in COCA showed a higher use of hyphenated compounds compared to *hydrogen*. *Hydro electric*, for example, showed just one occurrence, but 36 entries were counted for *hydro-electric*; *hydroelectric* amounted to 647 tokens.

4.2.2. An additional search was performed seeking the independent word *hydro-* in COCA. The results displayed in Table 8 confirm the autonomous character of this ICF.

Table 8. COCA. *Hydro-* entries and rate per million as a free ICF.

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	336	15	22	95	72	132
Per million	0.72	0.16	0.24	0.99	0.79	1.45

The analysis of the tokens found revealed a significant presence of *hydro* in academic prose from the field of science and technology. As *hydro* names a technical concept that refers to the English word ‘water’, the frequency rate reached in academic journals (39.28 %), highly over magazine (35.36%) and newspaper sources (21.42%) is unsurprising. Spoken registers recorded *hydro* far behind (4.46%). The free independent status of *hydro* can be observed in examples (5):

(5)

- a. In places such as the Pacific Northwest, integration of wind and **hydro** has become relatively common (ACAD. Mechanical Engineering, COCA 33, 2005)
- b. Solar, **hydro**, and tidal renewable energy sources are abundant, but environmental impacts and (ACAD. Bioscience, COCA 26, 2011)

As with *aero-*, a manually processing of the corpus discarded some instances of *hydro* as a free element, and attributed their bound morphological status in spite of the spaced spelling, as in: *hydro generators*, *hydro plant*, *hydro station* or *hydro turbines*. Ex.:

(6)

Two of the most successful projects described above embodied this principle, with Nepalese firms taking over the fabrication of **hydro turbines** and Indian firms that of wind turbine towers (ACAD. Environment, COCA 74, 1995).

4.2.3. Research from COCA revealed that the use of *hydro* has significantly decreased in the last decades; 2332 occurrences were obtained from the period 1990-1994 vs. 1559 from 2010-2012, thus exhibiting a 33% decrease.

4.3. Cyber-

Cyber- is defined in the *Online Etymology Dictionary*, as “an element in word formation, ultimately from *cybernetics* (q.v.). It enjoyed an explosive increase in usage with the rise of the Internet early 1990s. One researcher (Nagel) counted 104 words formed from it by 1994. *Cyberpunk* (by 1986) and *cyberspace* were among the earliest”.

4.3.1. A global frequency count of *cyber-* confirms its higher presence in magazine sources (29.26%), academic (24.42%) or newspapers (22.43%) than in spoken registers (14.43%) or fiction (9.44%), as indicated in the Table 9:

Table 9. *Cyber*: entries as a free ICF (COCA).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	5,847	844	552	1,711	1,428	1,312
Per million	12.60	8.83	6.10	17.91	15.57	14.41

The highest number of entries was documented in these words: *cyberspace* (1,638), *cyber* (947), *cybernetic* (171), *cyberpun* (141), *cybersecurity* (137), *cybernetics* (134), *cybersex* (119), *cyberbullying* (118), *cybercrime* (105), *cyberattacks* (59), *cybercom* (52), *cyberattack* (44). *Cyber* compounds used as adjectives ranked lower than nouns.

Taking *cyberspace* as a sample for *cyber-*compounds, COCA documented a comparatively significant use in magazines (31.68%):

Table 10. *Cyberspace*: entries and rate per million in different registers (COCA).

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	1,638	252	85	519	348	434
Per million	3.53	2.64	0.94	5.43	3.79	4.77

As graphically illustrated in Table 10, the higher use of *cyberspace* in magazines was followed by academic journals (26.49%) and newspapers (21.24%). *Cyberspace* in magazine journals nearly outnumbered spoken registers by as much as two to one (11.86%) and fiction by five to one (5.08%).

Throughout the corpus differences in spelling were observed. Instances of *cyber* spelled separately from the noun *space* were not present. The hyphenated format was found in 7 entries. Instances of solid compounds amounted to 1,638.

(7)

- a. *These include fine-grained systems (e.g., swarms of nano satellites), intelligent and autonomous vehicles, and very large-scale complex adaptive systems of systems, such as cyberspace* (ACAD. Mechanical Engineering, COCA 52, 2011)
- b. *Scientific, a global source of laboratory supplies and equipment, created Einsteinsgarage.com as a cyber-space warehouse and public auction site that buys and sells surplus, obsolete, and used scientific research, safety, and clinical supplies* (ACAD. Science, COCA 4, 2001)

4.3.2. The *Online Etymology Dictionary* postulates that *cyber*, as a stand-alone form, is attested by 1998 as short for *cybernetics* (which dates from 1948). *Cyber* entries from the corpus referred to computer or computer networks, the electronic medium in which online communication takes place. *Cyber* entries used independently without any additional inflectional ending amounted to 947 and reached the following figures in each register:

Table 11. COCA. *Cyber* entries and rate per million as a free ICF.

Section	ALL	Spoken	Fiction	Magazine	Newspaper	Academic
Frequency	947	275	70	198	254	150
Per million	2.04	2.88	0.77	2.07	2.77	1.65

9

These results demonstrate that the spoken registers rate the highest number of entries for *cyber* (29.03%), closely followed by newspapers occurrences (26.82%). Some examples exhibit the context in which *cyber* was used as a free morpheme:

(8)

- a. *He decided the best people to track down a bunch of cyber savvy students* (SPOK. NBC_RockCenter COCA 2, 2012)
- b. *School bullying among adolescents in the United States: physical, verbal, relational, and cyber* (ACAD. American Journal of Public Health COCA 66, 2012).

As for the grammar category of *cyber*, although Ralli (1992) had pinpointed the nominal category of ICFs, COCA documented entries belonging to other word categories: adjectival, verbal or even adverbial.

(9)

- a. *ADJ: Thus, although it is important to recognize when and if bullying is occurring, it is also vital to determine which forms the bullying takes (physical, verbal, cyber), as well as the content of the bullying* (ACAD. School Psychology, COCA 400, 2008)
- b. *VERB. Well, you're not supposed to cyber on work. But I'm going to try my best because I know there* (NEWS. CBS News Evening, COCA 215, 2011)

- c. *ADVERB: between how often these student groups reported being physically, verbally, socially, or cyber bullied* (ACAD. School Counsel, COCA 99, 2011)

4.3.3. The use of the ICF *cyber* exhibited a gradual increase from the period 1990-1994, where 475 entries were attested, to years 2012-2013 which exhibited 1,163 occurrences. A 145% increase was thus estimated.

5. Discussion

Regarding our first research question, on the differences shown by ICFs across registers, the results of the study of complex words with a neoclassical component have confirmed that the use of *aero-*, *hydro-* and *cyber-* varies according to the topic under discussion. A global account of statistical tables revealed the higher use of ICFs in magazines and academic journals, followed by newspapers. Spoken and fiction registers stayed globally far behind. The results evidenced that texts dealing with the fields of science and technology issues were the most prone to use *aero-* and *hydro-*, both as a constituent of compounds and as a free morpheme. However, a comparatively striking use of the independent morpheme *cyber* was attested in spoken registers.

The first hypothesis confirmed Evert and Lüdeling's (2001), and Baayen's (1994) postulate about the higher occurrence of neoclassical words in scientific texts. The assumption of the higher frequency of ICFs in magazines and academic journals was largely confirmed for *aero-* and *hydro-*. A detailed analysis of the magazine contexts in COCA where the ICFs were attested, revealed their scientific and technical character, a fact that supports the hypothesis set. Finally, the higher entries obtained for *cyber-* in spoken registers, evidenced its thorough transfer into everyday speech. The currency of the concept and its presence in daily life may account for its frequent use in spoken sources. This can be justified by the current undergoing process of getting a learned word introduced in today's daily life with the popularization of terminological units such as *cyber attack*, *cyber café*, *cyber chat* or *cyber Monday*. As for the spelling of ICFs, although in some instances, spaced compounds alternated with both concatenated and hyphenated compounds, the findings showed the tendency of frequently used compound to be more likely concatenated. COCA has shown the current trend of avoiding separate or hyphenated words in favor of solid words, mainly when the compound has adopted a special meaning and lexicalization has taken place.

As for the second research question on the free character of some ICFs, the data revealed that the lexical load of the ICFs provided *aero-*, *hydro-* and *cyber-* with full autonomous semantic content; this may be the result of a diachronic process of lexicalization, whereby an affix or a combining form becomes a free form and acquires independent lexical status. Specifically, *cyber* shows an increasing current use. Our hypothesis regarding the bound morpheme nature of ICFs, traditionally defined as stems that cannot be free, was not systematically supported by this research and was, thus, only partially confirmed. *Aero-*, *hydro-* and *cyber-* have nowadays developed into free morphemes and can now be used independently, a fact which contradicts the essence of the bound nature of combining forms supposed to be used only in combination with words, affixes, or other combining forms to create compounds.

Regarding the third research question on the variations expected on the frequency data of ICFs across registers in *COCA* over time, the hypothesis was partly confirmed. *Aero-* and *hydro-* showed a decreasing trend; they are possibly substituted by their equivalents *air* and *water* in science and technology domains. The outstanding increase of *cyber-* in the last decade manifests the novelty of the concept and the currency of the word-formation pattern. The use of *cyber-* is not exclusively today of higher, learned registers referred to science and technology. *Cyber-* has become part of the everyday vocabulary and may even have lost its Latin or Greek trace in language use. In this context, this study agrees with *The New York Magazine* (Dec. 23, 1996): "... nobody has an idea of the meaning of *cyber-* but it can be grafted onto any old word to make it seem new, cool and therefore strange, spooky" (cited in the *Online Etymology Dictionary*).

6. Conclusions and further research

This study has offered a tentative interpretation of a word forming process which, although not new, is still present today in the store of neoclassical lexical material currently in use in English. The controversy raised in its naming process has been addressed. The lexical load involved and their bound or free status and morphological categories have been discussed. *COCA* primed the present investigation into similarities and dissimilarities between registers based on the use of ICFs. *COCA*'s large textual basis facilitated a corpus-based analysis to obtain preliminary empirical evidence of the nature and extent of the lexical traits of ICFs adhered to different registers.

The results obtained shed some light on the current use of a sample of neoclassical morphemes in English across registers. The tendency of *areo-*, *hydro-* and *cyber-* to evolve from an initially bound morpheme into a free morpheme is an important contribution of the present study. The process of lexicalisation may be interpreted as the result of the dynamic characteristic of the language which exhibits a varying use of these combining forms across registers. On the basis of the findings, it can be stated that register variation exists in the language, and that variation exists across registers. The context of the situation in which language is used may affect the parameters of field, tenor and mode, proposed by Halliday (1980) and Halliday and Strevens (1964), making register as variety according to use, rather than function.

More continual research into language is required until a complete picture emerges. Ongoing research will further expand the sample and variety of ICFs to confirm the preliminary results here obtained on the tendency of certain ICFs to appear as stand-alone elements, an interesting finding here presented. Research into the native or nonnative etymology of the elements ICFs combine with can be the central topic of further studies. Further research may also imply a diachronic study of ICFs to document recent changes in the language, as important differences have been observed in *COCA* in the use of ICFs over time. Additionally, research can be extended to other corpora to see the incorporation of new terminological units related to the ICFs here studied. Since the corpora of the present study are made up of standard written texts (magazines, newspaper, fiction, academic journals) with few differences with each other in terms of register, "user-defined content" such as Web forums, blogs, social networks, etc., will be accounted for in future studies where substantial differences could probably be discovered.

The research here undertaken meets the current demand of the study of language in use, as neoclassical forms are a salient resource in the lexical expansion of terminology. From a pedagogical perspective, the different co-occurrence of ICFs in the various registers analyzed can

be kept in mind when defining the lexical contents of specific ESP textbooks. As English word-formation is usually taken for granted by teachers and planners, to account for how a text type can be characterized, the inclusion or exclusion of ICFs may reveal the degree of technicality or academicism of the text. Besides, pedagogical aspects of using corpus linguistics as an educational tool in higher education can be envisaged to introduce combining forms in the classroom as a trait of register variation.

References

- Adams, Valerie. 2001. *Complex Words in English*. Harlow: Pearson Education/Longman.
- Alvar, Manuel. 1993. *La formación de palabras en español*. Madrid: Arco Libros.
- Amiot, Dany; Georgette, Dal. 2007. Integrating neoclassical combining forms into a lexeme-based morphology. In Geert Booij, et al. (eds.), *On-line proceedings of the fifth Mediterranean morphology meeting (MMM5)* Fréjus 15-18 September 2005, University of Bologna.
- Baayen R. Harald. 1994. Derivational productivity and text typology. *Journal of Quantitative Linguistics* 1:6-34.
- Bauer, Laurie. 1998. Is there a class of neoclassical compounds, and if so, is it productive? *Linguistics* 36(3):403-422.
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus linguistics and linguistic theory*, 8(1):9-37.
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Corbin, Danielle. 2001. Préfixes et suffixes: du sens aux catégories. *Journal of French Language Studies* 11(1):41-69.
- Davies, Mark. 2008(-). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- Díaz-Negrillo, Ana. 2014. Neoclassical compounds and final combining forms in English. *Linguistik Online*, 68(6).
- Ferguson, Charles A. 1982. Simplified registers and linguistic theory. In *Exceptional Language and Linguistics*, Loraine Obler and Lise Menn eds., 49-66. New York: Academic Press.
- Fischer, Roswitha. 1988. *Lexical Change in Present-Day English: A Corpus-Based Study of the motivation, institutionalization and productivity of creative neologism*. Tübingen: Gunter Narr Verlag.
- Evert, Stefan; Anke Lüdeling. 2001. Measuring morphological productivity: Is automatic preprocessing sufficient? In *Proceedings of the Corpus Linguistics 2001 Conference*, Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie and Sreen Khoja (eds.), 167-175. Lancaster: Lancaster University.
- Iacobini, Claudio. 2004. Composizione con elementi neoclassici. In *La formazione delle parole in italiano*. Maria Grossmann; Franz Rainer (eds.), 69-95, Tübingen: Max Niemeyer Verlag.
- Kastovsky, Dieter. 2009. 'Astronaut, astrology, astrophysics': About Combining Forms, Classical Compounds and Affixoids. In *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX2)*, Roderick W. McConchie, Alpo Honkapohja; Jukka Tyrkkö (eds.), 1-13. Somerville, MA: Cascadilla Proceedings Project.
- Halliday, Michael. A. K. 1980. Register variation and identity of a text. *Sophia Linguistica* 6:60-79.

- Halliday, Michael A. K.; Peter Strevens. 1964. *The Linguistic Sciences and Language Teaching*. London: Longman.
- Lüdeling, Anke. 2006. Neoclassical word-formation. In *Encyclopedia of Language and Linguistics*. Keith Brown (ed.), 217-227. Oxford: Elsevier.
- Lüdeling, Anke; Tanja Schmid; Sawwas Kiokpasoglou. 2002. Neoclassical word formation in German. *Yearbook of Morphology* 2001:253-283.
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation*. München: C.H. Beck.
- Martinet, André. 1979. *Elementos de lingüística general*. Madrid: Gredos.
- McCauley, Jane. 2006. Technical combining forms in the third edition of the OED: word-formation in a historical dictionary. In *Selected Proceedings of the 2005 Symposium on New Approaches in English Historical Lexis (HEL-LEX)*, Roderick W. McConchie et al.(eds.), 95-104. Somerville, MA: Cascadilla Proceedings Project.
- Montermini, Fabio, 2002. *Le système préfixal en italien contemporain*. Unpublished doctoral dissertation. Université de Paris, Nanterre (France) and University of Bologna.
- Oxford English Dictionary* (3rded). <http://public.oed.com/how-to-use-the-oed/what-is-the-oed-online/>
- Online Etymology Dictionary*. Available: <http://www.etymonline.com/index.php>
- Petropoulou, Evanthia. 2009a. Neoclassical Compounds in Modern Greek and English: A comparative view. *Theoretical and Applied Linguistics*, 87: 67-98.
- Petropoulou, Evanthia. 2009b. On the parallel between neoclassical compounds in English and Modern Greek. *Patras Working Papers in Linguistics*, 1, pp-40
- Prčić, Tvrtko. 2005. Prefixes vs. initial combining forms in English: A lexicographic perspective. *International Journal of Lexicography* 18: 313–334.
- Prčić, Tvrtko. 2008. Suffixes vs final combining forms in English: A lexicographic perspective. *International Journal of Lexicography*, 21(1), 1-22.
- Ralli, Angela. 2008. Greek deverbal compounds with ‘bound stems’. *Southern Journal of Linguistics* 29(1-2):150-173.
- Ralli, Angela. 1992. Compounding in Modern Greek. *Rivista di Linguistica* 4(1):143-174.
- Ralli, Angela. 1988. *Eléments de la morphologie du grec moderne. La structure du verbe*. Doctoral Dissertation. Université de Montréal.
- Scalise, Sergio. 1986. *Generative morphology*. Dordrecht: Foris.
- San Evaristo, Eva. 2010. El uso de los compuestos clásicos en bioinformática. *LSP. Professional Communication, Knowledge Management and Cognition* 1(1): 29-46.
- Selkirk, Elisabeth. 1982. *The syntax of words*. Cambridge, MA: MIT Press.
- The American Heritage Dictionary of the English Language*. 2009. Fourth Edition. Houghton Mifflin Company [2000].
- Scher, Ana Paula; Vitor Augusto Nóbrega,. 2015. Unifying neoclassical and stem-based compounds: a non-lexical approach. *Revista LinguiStica*, 10(1).
- Williams, Edwin 1981. On the notions “Lexically related” and “Head of a word”. *Linguistic Inquiry* 12.2: 245–74.
- Zwicky, Arnold M. Z.; Ann D. Zwicky. 1982. Register as a dimension of linguistic variation. In Richard Kittredge and John Lehrberger (eds.), *Sublanguage. Studies of Language in Restricted Semantic Domains* (27): 213-8.

APPENDIX.

Abbreviations used:

ACAD: Academic

COCA: Corpus of Contemporary American English

ESP: English for Specific Purposes

FIC: Fiction

ICF: Initial Combining Form

MAG: Magazines

NEWS: Newspaper

SPOK: Spoken