LEGE ARTIS
Language yesterday, today, tomorrow
Vol. II. No 2 2017

THE DIACHRONIC DEVELOPMENT OF COMBINING FORMS IN SCIENTIFIC WRITING

Katrin Menzel[1]*, Stefania Degaetano-Ortlieb

Corresponding author*

Menzel, K. & Degaetano-Ortlieb, S. The diachronic development of combining forms in scientific writing. In Lege artis. Language yesterday, today, tomorrow. The Journal of University of SS Cyril and Methodius in Trnava. Warsaw: De Gruyter Open, 2017, vol. (2), December 2017, p. 185-249. DOI: 10.1515/lart-2017-0016 ISSN 2453-8035

Abstract: This paper addresses the diachronic development of combining forms in English scientific texts over approximately 350 years, from the early stages of the first scholarly journals that were published in English to contemporary English scientific publications. In this paper a critical discussion of the category of combining forms is presented and a case study introduced to examine the role of selected combining forms in two diachronic English corpora.

Key words: combining forms, morphology, history of scientific English, language for specific purposes, information density, corpus linguistics

1. Introduction

In this paper, we examine the diachronic development of combining forms (henceforth CFs) in English scientific texts ranging from the first scholarly journals published in English at the end of the Early Modern English period to contemporary English scientific publications. Our case study has a particular focus on combining forms from the Graeco-Latin stock of lexical morphemes as we assume this to be a productive

[1] This study was funded by the German Research Foundation (DFG) in the framework of the project 'Information Density and Scientific Literacy in English: Synchronic and Diachronic Perspectives' in the Collaborative Research Center (SFB 1102) with the title 'Information Density and Linguistic Encoding' (http://www.sfb1102.uni-saarland.de) and EXC 284: Multimodal Computing and Interaction (www.mmci.uni-saarland.de). We also would like to express our gratitude to Stefan Fischer for his help with preparing the datasets for the analyses.

resource for one of the major word formation processes in English for specific purposes (ESP) from the 17th century onwards. In particular, we consider the neoclassical combining form-lysis. Combining several lexical morphemes within a single lexical item is a word formation strategy that is particularly important for informational texts from scientific and technical domains. This word formation process helps to avoid longer alternative constructions such as multi-word terms or phrasal structures. For analysing the evolution of English scientific discourse with regard to the role of CFs, we use two diachronic corpora of scientific texts covering various disciplines and ranging from the middle of the 17th century onwards to the beginning of the 21st century: the Royal Society Corpus (RSC) (Kermes et al. 2016) and the Scientific Text Corpus (SciTex) (Degaetano-Ortlieb et al. 2013). In the analysis we test the following hypotheses: (H1) Conventionalized use of CFs over time, i.e. we test whether CFs are increasingly used with the same stems or combined with a variation of stems over time, and (H2) Interaction between convention and productivity, i.e. we test whether the use of CFs becomes more conventionalized and whether these forms become more easily available and productive in different, yet closely related, analogous grammatical contexts (cf. De Smet 2016).

In terms of methods, our approach is different from traditional approaches that only focus on observed frequencies of certain morphemes or type-token ratios of words as they occur within diachronic corpus data as an indicator of morphological productivity. What we primarily consider is the surprisal value of each unit, i.e. the probability of a unit occurring in a given context (Hale 2001; Levy 2008). Thus, rather than using unconditioned frequencies, we use conditioned probabilities (see also Degaetano-Ortlieb & Teich 2016 for a comparison of surprisal and type-token ratio to investigate productivity), i.e. the probability of a CF occurring with a particular stem. We present how the notion of surprisal leads to some insights on the diachronic development of CFs and the elements with which they co-occur in complex lexemes alongside a possible change in their grammatical properties over time.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

After this introductory section, Section 2 will provide a thorough background section and literature review to capture the heterogeneity of approaches and fields of linguistics to which the concept of CFs is relevant, as well as the complexity of issues that continue to play a significant role in linguistic discussions of CFs. We will discuss several factors that have contributed to the fact that the category of combining forms has a rather broad reference in contemporary linguistic publications. Section 3 will specifically address the role that neoclassical combining forms, derived from nouns or verbs in classical languages, play in English scientific writing. Section 4 will explain our hypotheses, methodology and data and present our case study of one neoclassical combining form in a diachronic corpus of English scientific texts. In Section 5, we present our conclusions.

2. The category of combining forms

2.1 Combining forms in lexicographic and didactic resources

The labelling of certain word-forming elements as 'combining forms' is a relatively recent practice that has received some scholarly recognition in the morphological and lexicological literature, but the exact status of CFs has become the subject of some critical discussion since the term was introduced. This subsection will focus on the use of the term 'combining forms' in practical-oriented lexicographic and didactic resources as they play a prominent role in shaping the understanding of what a CF is, while Section 2.2 will summarize how lexicologists and morphologists have attempted to characterize and clarify the concept from a theoretical and methodological perspective. It is outside the scope of this paper to discuss in depth all forms that fall under this term, but we will nevertheless address a range of aspects from current discussions on various types of combining forms. For this paper, we will then narrow the concept down to prototypical forms of a very specific type to be used in our corpus analysis.

Not all English dictionaries and didactic materials with a certain amount of information on word formation processes or word-internal structures use the term 'combining

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

forms'. Some prefer not to make very fine-grained distinctions between different types of word beginnings and endings. Nevertheless, various prominent monolingual English dictionaries and some didactic resources for ESL/EFL for specific purposes and for English as a foreign language (EFL) apply the category of combining forms to initial and final bound lexical elements in the classification and description of complex words and their components. The most prominent dictionary that contributed to the adoption of the term 'combining forms' is the Oxford English Dictionary (OED). In fact, the term goes back to the predecessor of the OED, the New English Dictionary, which was published in several volumes between 1884 and 1928 (cf. Kastovsky 2009a: 2), at a time when modern morphological and word formation descriptions for English were still in their formative stages. The same term was also used in the second edition of the OED (1989) and is still applied in the classification of dictionary entries and their components in the third edition of the OED (in progress; the OED is currently revised and updated online four times a year). The structure of complex words is briefly described in the generic headword section of respective dictionary entries. Additionally, the OED generally provides some morphological and grammatical information in the etymological section of the entries, noting the process of derivation and how a word was formed when it entered the English language. Several types of word parts that occur as bound elements within English lexemes (prefixes, suffixes and CFs) fall under 'special types of main entries' of the OED, and a selection of these elements has separate entries.

There are currently 2275 entries classified as combining forms in the online OED, which can be queried via the Advanced Search page of the OED as a type of 'Part of speech'. Examples of CFs in the OED are Anglo- in Anglo-Irish, bio- in biology, or -lysis in electrolysis. The origin and history of the individual CFs are briefly discussed for each of these word-forming elements in the OED. As a group of similar morphemes with lexeme-like semantics, but some formal properties of affixes, they play an important role in many scientific formations in combination with other

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

elements that are also often of Latin or Greek origin (i.e., other combining forms, affixes or free bases and lexemes, e.g., combining form *photo-* + combining form *-graphy* in *photography*, *phot(o)-* + suffix *-ic* in *photic* or *photo-* + lexeme *effect* in *photoeffect*). They also play a certain role in general, non-technical language and various types of lexeme formation processes, for instance in blending or clipping (cf. Section 2.2).

The combining forms in the online OED can be displayed as a list, sorted alphabetically or by their date of first use in the OED citation corpus. It is also possible to query only those CFs with a certain language of origin or those forms whose entries start or end with a hyphen if one wants to distinguish between entries for initial and final combining forms. Some forms can be used in both positions, but the number of elements typically used as initial combining forms in the OED is approximately four times higher than that of final combining forms. This is due to the fact that initial combining forms listed in the OED in many cases have a more specific lexical meaning, while final combining forms are often derived from more generic lexemes and can be combined as heads of complex nouns with initial combining forms, e.g., the final combining form *-graphy* with the literal meaning ('writing') or a less literal meaning ('description', 'recording' or 'field of study') in *astrography*, *calligraphy*, *cryptography*, *geography*, *photography*, *stenography*, etc. Although these words may involve recognizable morphemes, the sender who selects the elements of a message and the receiver to whom a text is addressed do not necessarily parse such internally complex words into their constituent morphemes when they use them. Morphological awareness, the skill to analyse internal structures of complex words and to understand morphological rules of the native language, is a comprehension and language production skill that has to be acquired by language users, along with other linguistic skills. Numerous CFs in the online OED can be traced back to Greek or Latin content words. Most entries for words in the OED involving at least one combining form seem to be neoclassical coinages, but some have also been borrowed directly as compounds from classical languages (e.g., *bibliography*

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

that goes back to the classic Greek compound bibliographia ! . ¥ . 6 R P H were coined in scientific English or post-classical scientific Latin or borrowed from other European languages such as French or German (e.g. gpt(o)- + gram, after German Optogramm).

The distinction between prefixes and suffixes and CFs as well as the boundary between derivation and compounding may not always seem consistent in English dictionaries. There are far more entries for combining forms than for affixes in the OED, but being a potentially open-class category due to their lexeme-like semantics, it may seem astonishing that less than 2300 elements are identified as combining forms. These seem to represent only a selection of CFs that have been used or are still used productively in word formation processes in English. On the one hand, the actual number of final combining forms in the list may be even smaller than the query results indicate because the OED lists final elements such as -graph, -grapher, -graphic, -graphical, and -ography as separate items that all involve the same root morpheme or with additional suffixes and / or linking elements attached. English has acquired some CFs as doublets or allomorphs that may have separate dictionary entries (dent-, dento-, and odonto- as adaptations from the Latin and Greek words for 'tooth'; histo- with Latin origins and historico- from Greek). There are also some elements derived from suffixes without lexical meaning in classical languages that have also been listed as combining forms in the OED (e.g. -ene in chemical terms such as benzene or naphthalene). On the other hand, it would be possible to identify more CFs in English in addition to those in the OED to which the label has been assigned. In graphology, for instance, the initial element is identified as a combining form in the etymological note for this word, but there is no separate entry for graph(o)- as initial combining form in the dictionary in contrast to the above-mentioned entries for this morpheme in final position. Additionally, there is also a free morpheme graph in English that can occur on its own. It has three different entries in the OED with different etymological notes i. shortened from graphic formula, ii. derived directly from the Greek word for writing or

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

and iii. shortened via clipping from words that have -graph as a final element such as chromograph (from chromo-, comb. form of chromium, a Latinized form of the French chrome, ultimately from the Greek word for 'color'). Some elements with lexical meaning such as -some in chromosome, lysosome etc. from Greek σῶμα 'body' have been listed as suffixes.

The lists of CFs in dictionaries can only give us a rough estimation of how many of these forms actually exist in the English language. Dictionary entries in the OED and other resources for less common words such as biblicide (i.e. the 'killing' of a book, derived from two Latin words) do not label the initial element, in this case (i)bibli-, as a combining form, as it does not occur in many other lexemes in combination with a final CF, but a more common element -cide is listed as a combining form, while for instance, in biblioklept (i.e. a book-thief, derived from the Greek words for 'book' and 'thief') only the initial element has been assigned the status of a combining form in the OED. Klepto- has an entry as a combining form, but only as an initial element in words that also have final CFs and involve the -o- as a linking element such as kleptocracy or kleptomania. It also has a separate entry as a slang expression klepto as a reduced form of kleptomaniac, and there are two entries for the rarely used adjectives kleptic and kleptistic in which the initial element is combined with a suffix, but not described as a combining form in the etymological note. Less common words generally tend to have short etymological sections in the dictionary. Various classical elements may have been initially introduced as unique or rare morphemes in borrowed complex words. If their status as a part of a word is transparent to native speakers, with the result that their lexical meaning can be recognized, they might be analysed as CFs at a later stage when they become more productive as bound elements in other words or even as free morphemes.

As technology and science permeate nearly all areas of life in modern times, there is a certain trend for standard dictionaries to bolster their technology vocabulary and to

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

identify more components, and hence also more CFs, in technical terms. On the other hand, various new coinages with borrowed or native combining forms have made their way into standard dictionaries yet as such words may function as occasionalisms whose usage is limited to certain contexts, as in scientific or technical discourse. According to Haspelmath (2002: 216), productive morphological rules are likely to produce numerous occasionalisms, but new words formed by such rules are sometimes hardly noticed consciously by speakers, hearers, and lexicographers as they may not strike them as particularly innovative. Other words with combining forms may be subject to fashion or regional preferences and do not enter the dictionary for those reasons.

The group of combining forms in the OED can be visualized on a timeline to show approximately when these forms entered the English language. Most items classified as CFs have their first citation in the OED quotation database between 1800 and 1899, but they may already have been in use slightly earlier, at least in specialized registers. For most combining forms themselves, no frequency information is given in the OED, but all entries for lexemes, apart from obsolete items, have been labelled with some frequency information (for written present-day English, derived primarily from Google Books Ngrams data, cf. the section "Key to Frequency" in the Online OED). Frequency bands in the OED run from 8 for very high-frequency words to 1 for very low-frequency words. Among the most frequent words with CFs, for instance, those ending in -graph, we find lexical items such as paragraph and photograph which have been assigned to Band 6. In Band 6, there are words that occur between 10 and 100 times per million words in modern English usage (such as many nouns referring to specific objects or processes). Band 1 contains almost 20% of all obsolete OED entries, which are often highly technical, but archaic or non-standard terms and which would be very rare in modern texts from well-balanced, large corpora. An example of a Band 1 word ending in -graph is selenographi, i.e. a photograph of a part of the surface of the moon, a word for which we find a citation in the OED from an academic article

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

published in the Proceedings of the Royal Society of London in the late 19th century. As many words with similar structural features are rare in average texts or general corpora of modern English, they are sometimes assumed to play only a marginal role in the English language as a whole and to be rather unproductive, apart from particular registers such as scientific or technical English where they can occur in terminology. However, in diachronic English corpora, such as our specialized corpora of English academic writing, they are related to important register-specific word formation patterns and they are also very interesting from a cross-linguistic perspective.

What falls under the definition of 'technical combining forms' in the OED is explained in McCauley (2006). The OED considers initial and terminal elements as subtypes of these forms. A short definition can also be found in the entry for combining forms on the website of Oxford Dictionaries. In this entry, combining forms are illustrated with various examples and broadly defined as forms of words normally used in compounds in combination with another element and as elements that contribute to the particular sense of words. In the section "Guide to the third edition of the OED" on the OED website, combining forms are described as "words which occur in a slightly altered form when used to introduce longer compound words (such as 'medico-' for 'medical')". Other prominent dictionaries that use the term also give a rather brief and relatively broad definition and illustrate the category with a selection of native and non-native morphemes of various origins and initial and final bound stems as well as altered word forms that occur only in compounds.

The Merriam Webster describes a combining form as a "form of a word that only appears as part of another word" and adds that "[c]ombining forms are similar to affixes but can have a bit more lexical substance to them". Their subtypes in the dictionary are classified according to the type of word classes in which the form can be used. The examples given are final combining forms such as -graph in photograph that falls under the subtype of a 'noun combining form'; -lyze in electrolyze, a 'verb combining form',

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

or -wise in clockwise, which is called an 'adverb combining form'. In the Macmillan Dictionary, a combining form is "a form of a word that has its own meaning but is used only in combination with other words to make new words, for example -footed in 'a four-footed animal'". These examples demonstrate that dictionaries using the term usually tend to give a relatively broad definition and illustrate the category of combining forms by various elements that play a certain role in English word formation processes and that can neither be clearly identified as affixes nor as independent words due to their semantic and formal properties.

Nowadays lexicographers tend to include native morphemes of various origins in the list of combining forms if they occur only in combination with other elements, but not as independent words in English (e.g., Greek -(o)polis, French -ville, German -meister). Native morphemes with adjectival meaning that occur in combination with other elements as bound terminal elements with a specific sense (e.g., -like in birdlike or -wise in clockwise) seem to be another relatively recent addition of lexicographers to the category of combining forms. Recently, some truncated words that have undergone clipping (e.g., burger shortened from the word hamburger, -gate in the sense of 'scandal' from the word Watergate), but also pseudo morphemes with classic or Romance origins that have undergone semantic and structural reanalysis by processes of analogy (e.g., -holic in coinages such as workaholic or -(a)-thon in edit-a-thon in analogy with the model words alcoholic and marathon) have equally been subsumed under the category of CFs by various scholars and in lexicographical resources. Such elements are a productive source for novel blends in creative language use, media discourse and quasi-technical jargon. Recent word formations such as the above-mentioned edit-a-thon are often not (or not yet) listed in standard dictionaries, but the OED nowadays lists -thon as a suffix and at the same time as a variant -athon, which it classifies as a combining form. Historically there is no morpheme -thon/-athon. The shortened form acquired a new meaning associated with the sense of the entire original word. The OED etymology section notes

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

that it has been 'barbarously' extracted from the word marathon and that it originally was found in occasional American coinages, but rarely in Britain, denoting something carried on for an abnormal length of time. It can now more freely be combined in English neologisms that copy stress patterns or syllabic structures of neoclassical formations phonologically. These examples demonstrate the blurred distinction between affixes, CFs and other parts of words. The above-mentioned form -gate, for instance is assigned the status of a combining form in the respective OED entry, but in texts written by OED editors targeting the general public, it is usually called a suffix (e.g., Maier n.d.). In an interview with The Guardian on neologisms and ongoing language change, Maier also used the term 'suffix' for this element and compares its behavior to final elements in blends such as -ther 'suffix' in the word Brexit (Kean 2017). Clipped word fragments as in 'Brexit' have been called 'splinters' or 'fracto-lexemes' in the theoretical literature (Bauer et al. 2013: 522–530) if they occur in lexical blends. Such terms are not used widely in lexicographical resources, introductory linguistic textbooks, and media texts on linguistic topics that prefer to present information in a manner which is as audience-friendly as possible, avoiding an excessively high complexity of distinctions and technical linguistic terminology. Nevertheless, lexical blending as a process of lexical creativity attracts some interest of the media and is entering public consciousness via recent buzzwords from media and advertisements texts that trigger further wordplay.

Lexicographers have repeatedly revised and adapted their definitions and lists of CFs and updated the number of entries that fall under the category. We refer the interested reader, for instance, to the discussion on the development with regard to the treatment of combining forms in the OED explained in Durkin (1999: 29–32). In the early editions of the OED, the documentation of CFs still proved more difficult than it is nowadays as they can occur as parts of different lexical categories. They neither share any specific semantic features nor can they be identified through a certain length or combinations of letters. Particularly, the systematic identification of final combining forms was not

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

an easy task, as alphabetically ordered lists could not be sorted easily in different ways to identify and document words with certain internal structures or components systematically, which has become less difficult with the advent of digitization in lexicography. Additionally, more insights from contrastive studies about similar types of CFs in other European languages have been woven into English resources in the last decades, while in the early stages of English dictionary compilation no extensive comparisons with similar morphological structures in other languages were made. In the last decades various cross-linguistic studies have contributed to the identification of cognates in other languages. Lexicographical resources and studies in Romance languages in particular have contributed to the current understanding of CFs in English, which means that the provision of coinage information and combining form entries in the online OED is in numerous cases analogous to the information in etymological notes in the resources VXFK DV WKH 7UpVRU GH OD ODQJXH IU closest French counterpart of the OED covering several centuries of use. What falls under CFs in English is most frequently GHVFULEHG DV pOpPHQW IRU French resource, although it sometimes uses different terms pOpPHQW SUp pOpPHQW IRUPDQW DQG ±for elements with a similar function, VLW e.g., agro-, *Américano*-, bio-, bibli(o)- as prefix elements, but -grapho-, Italo- or lys(i)-/lys(o)- HWF DV pOpPHQWV DYHF XQH IRQFWLRQ VLPLODLUH composition'.

While some items that are identified as initial combining forms in lexicographic resources end in a vowel that is seen as a part of that morpheme in the respective entries (e.g., Anglo-, Graeco-), others are analysed as consisting of two parts: the combining form itself and a separate linking element, or a 'connective' as it is called in the OED (most frequently the vowel -o-, but also sometimes -a- or -i-), that is attached to non-native items when they are combined with other elements, e.g., music(o)-, lyric(o)-. Music and lyric also occur as etymologically related free lexemes of the respective combining forms, but the bound and free forms exhibit different, register-specific

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

distribution patterns in the language. In some elements -o- was part of the Ancient Greek stem, e.g., in *opto-* IURP QŒ2)" YLVLEOH DQG FDQ EH the form that is listed as a combining form in the OED, but is also sometimes analysed as a connective or linking vowel in the respective dictionary entries. Bauer (1998), 3UüLü and Kastovsky (2009a: 6), among others, discussed the status of connective -o- and other linking vowels in neoclassical compounds without showing a strong preference for one specific description and presenting several options for analysis. Such elements can be regarded as separate linking elements, e.g., -o- between *phot* and *graph*, or as a part of the first or the final element, or as being a part of both the first and the final element at the same time. Hamans (2014) argues that there are allomorphs that co-exist in the English language (e.g., *-graph* and *-ograph*).

Dictionaries vary to a certain extent with regard to which morphemes they include in the list of CFs. They are typically revised regularly to eliminate inconsistencies within themselves (e.g., the entry for *opto-* is now identified as a combining form in the OED, but in previous editions it was not labelled with any specific part of speech, allowing potential ambiguity in whether it should be seen as a combining form or a prefix (McCauley 2006)). Productivity and transparency for the intended user group and aspects of language change seem to play a role for lexicographers when they decide to assign word formation elements to a specific category. The fact that the tasks of lexicographers are more practically oriented than those of lexicologists and morphologists may serve another reason for some inconsistencies at the theoretical level in the application of morphological categories in the dictionary entries. The debate on how much grammatical and morphological information, and particularly how much information on specific morphemes as 'problem zones' in word formation theory such as CFs, should be included in monolingual and bilingual dictionaries and dictionaries of languages for special purposes (Elsen 2013a; Grimm 1997; Mugdan 1989), ended in no definite overall conclusion.

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Empirical research on dictionary use, including surveys to identify the needs, H[SHFWDWLRQV DQG EHKDYLRU RI GLFWLRQDU\ X V 2013: 197). However, for the majority of users of standard, learner's or technical dictionaries ± in contrast to the linguistic scholar ± it is probably of minor importance whether morphemes such as -(o)polis or -(o)logy are labelled as combining forms, suffixes or just as final elements of words in the respective entries or whether, for instance, the -o- should be analysed as either a separate connective or as part of such an element. What most users would expect from a dictionary is primarily that its entries help to unlock the meaning of unfamiliar words and their components. Current lexicographical studies and research into dictionary use underline the central role of the needs and expectations of the users and emphasize the function of dictionaries as a useful tool for their intended audience in certain situations (cf. Spitzer 2016: 293ff). In existing resources, lexicographers have often decided to "gloss over certain distinctions that the theory-minded lexicologist would want to introduce" (Kastovsky 2009a: 1). Although there is potentially enough space for detailed morphological information in the entries of modern online dictionaries, editors of such dictionaries do not intend to give lengthy descriptions and very detailed morphological and etymological information in the entries unless the resource was compiled particularly for that purpose.

Apart from the relatively brief information on the morphological status of word parts that some large and contemporary general dictionaries such as the Online OED, Merriam-Webster Online or Macmillan Dictionary Online provide for their entries, there are also some English terminology dictionaries and resources on English complex words derived from Latin and Greek elements that make some basic distinctions between different types of word formation elements in their entries (e.g., Ayers 1965, 1972). A few specific English dictionaries, thesauri and other types of didactic resources of different sizes and with different selections of thematic categories typically try to avoid making a clear distinction between word roots and combining

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

forms, on the one hand, and combining forms and affixes, on the other hand (cf. for instance the resources edited by Borror 1960; Danner 2014; Denning et al. 2007; Quinion 2003; Robertson 1991; Sheehan 2000; Smith 1969; Urdang 1982, 1984, 1986). In these resources, combining forms and other elements are typically presented as one group and labelled with more general or superordinate terms in their description such as 'word parts', 'vocabulary elements', 'word beginnings and endings', 'initial and terminal elements' or 'word-initial and word-final elements'. In fact, the term 'combining form' also entails a certain degree of unspecificity. Items with that label could basically have any form, origin, or semantic content and occur in various positions within lexemes in which several elements have been combined. Didactic resources discussing the etymology and construction of scientific terms with the intention of helping students and professionals understand and remember the terminology of their fields sometimes use the term 'combining form' but do not apply it in any strict sense. A textbook on dental terminology for instance has a section on the structure of complex terms and contrasts prefixes and suffixes with roots used as CFs (Dofka 2013: 3–15). However, elements added to the end of root words or to the end of CFs are generally labelled as suffixes in that book, e.g. -gram or -graphy, which does not conform to current conventions of either avoiding the term 'combining form' entirely or analysing neoclassical compounds as consisting of more than one combining form.

What we are not able to conclude with certainty from the information available on various CFs in lexicographical resources is the actual total number of this group of morphemes in English, how productive these elements are and how many different words have been coined with them in English as a whole or in particular time spans or registers. For this reason, it is useful to complement the information obtained from these resources with specific types of corpus analyses as we outline in this paper.

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

This section has aimed to discuss some aspects of combining forms in lexicographic and practically-oriented resources, while the next section will provide an overview on more theoretically-oriented English and cross-linguistic studies and on recent productivity studies and publications on word formation processes involving CFs.

2.2 The status of combining forms in lexicological and morphological studies

The boundaries between theoretical accounts of morphology and applied linguistics or between research and textbook-like accounts of word formations written for pedagogical purposes are not always clear-cut in the existing literature. Moreover, some theoretical aspects are fundamentally interwoven with issues that have already been addressed in the previous section on the treatment of combining forms in lexicographical resources. This section will add a few remarks on the status of CFs as discussed in existing handbooks, journal articles and other publications on English morphology and lexicology.

The applications of the term 'combining forms' in lexicographical resources attracted some criticism from linguistic scholars such as Marchand (1969: 131-133) due to some remaining inconsistencies. Several scholars have worked on the interface between lexicography, lexicology and morphology. As outlined in the previous section, various linguistic publications, such as the above-mentioned studies by Mugdan (1989), Grimm (1997) and Elsen (2013a), raised the questions of how specific the linguistic information on CFs and related types of morphemes should be in didactic materials, what aspects from the theoretical literature on morphology and word formation theory should be reflected in lexicographic resources, and what the practical consequences of different approaches would be. Others have published articles from both a linguistic and a lexicographic perspective on the structural analysis of bound formatives in complex English words or the borderline between compounding and derivation, which is difficult to draw, especially when CFs are taken into consideration (e.g., Ha F N H Q + D F N H Q 3 D Q R F; 2005; 2007; 2008; Ü L Ü

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Diachronically, many pre- and suffixes may have developed out of compounding processes, but they can typically be traced back to closed- class morphemes or have adverbial or prepositional counterparts. Borrowed English pre- and suffixes with Latinate or Greek origin often occur only as bound morphemes in English and not as independent lexemes, similarly to borrowed CFs, but they fall into a smaller number of semantic classes. Borrowed prefixes, for instance, can premodify their bases expressing adverbial-like meanings such as quantity or number (e.g., bi-, poly-), direction, location or temporality (ab-, ante-, de-), degree or size (hyper-, micro-, ultra-), and negation (anti-, dis-, non-). Suffixes can be distinguished from final combining forms due to their strong functional and grammatical character, e.g., a category-determining function or denoting abstract and general concepts such as actions, quality or state (e.g., acid + suffix -ity).

In the morphological and lexicological literature, the exact status of neoclassical and native combining forms has been the subject of some recent critical discussion as these types of bound morphemes that are found not only in English, but in other languages as well, are difficult to define precisely in operationalizable terms. Due to terminological disagreements among scholars, some have even suggested that the notion of combining forms should be abandoned altogether by arguing that other morphological categories such as 'stems' are sufficient to describe the same type of word formation processes, e.g., stem compounding instead of compounds based on CFs. Kastovsky (2009a: 12) suggested a scale of prototypical patterns of word formation processes arranged from independent towards less independent constituents, i.e., compounding > stem compounding > affixoids > affixation proper > clipping compounds > blending > splinters > acronyms.

Combining forms are sometimes also referred to as confixes (Ke), particularly in the Germanophone academic community, e.g., Donalies 2009; Elsen 2005;)OHLVFKHU 0LFKHO 2013. The interested reader is referred to the

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

discussion of compounds with bound stems and confix compounds in Eins 2012: 307ff and his mild criticism (ibid: 311) of the 'confix boom'. The rapidly growing interest in the topic led to the fast establishment of the confix term in word formation theory, which contributed to the fact that it acquired the status of a central unit of German morphology with a rather broad definition. A similar development took place with regard to other languages. Non-native combining forms in English as a part of neoclassical compounds closely resemble the concept that is discussed under the term R I N R P S R V L W L R Q J (confix that can be used in compounds). It has been pointed out that the terms confix and combining form in general are sometimes used synonymously (e.g., Zelle 2016: 11), while others have pointed out differences E H W Z H H Q W K H V H F R Q F H S W V 5 D G L P V N ê lexical meaning 'semi-words' in his description of neoclassical compounding patterns in Italian. Combining forms have also been labelled as (or have a certain amount of overlap with) affixoids, quasi-lexemes (Warren 1990), semi-prefixes and semi-suffixes, lexical affixes, affix-U H V H P E O L Q J H O H P H Q W V 3 U ü L ü have subsumed various elements from different languages under these headings that can neither be clearly categorized as affixes nor independent words due to their lexeme like semantics and formal properties of affixes. In our opinion, the notion of CFs has not yet been sufficiently developed and has not always been applied consistently in the theoretical literature. Some publications have adopted a rather broad view of what falls under the term of combining forms or leave it relatively open where the boundaries of this concept lie, so that it is only partially operationalizable in empirical studies on the class of combining forms as a whole. This perspective may involve ranked prototypicality scales as in Seiffert (2008: 103) and no clear cut-off point to transitional phenomena or a division of CFs into several semantic, etymological or structural subgroups, as for instance Warren (1990) does by suggesting that Group I forms represent allomorphs of source words and Group II and Group III forms are different types of truncated forms or parts of model words. The risk with a broad definition is that the term 'combining forms' might become a 'quicksand term' with a very general

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

meaning, i.e. a term that has so many conflicting definitions and connotations that it leads people into a conceptual quicksand (Nord & Connell 2011). It might be worthwhile now to tighten up terminology and to narrow down definitions to reduce existing ambiguities and in order to facilitate empirical analysis.

The problematic nature of the term has frequently been addressed, but it has by no means been resolved, and there are far more monographs and other types of academic publications on more prominent word formation elements. Linguists that observe English word formation processes from either a synchronic or diachronic perspective or both (e.g., Bauer 1983: 26; Stein 1973, 1977) have recognized the fluid boundaries between processes involving CFs such as neoclassical compounding and other types of word formation such as blending, compounding, affixation, clipping and forming acronyms. The word formation processes involved in the creation of technical or jargon terms containing combining forms are typically addressed rather briefly in contemporary handbooks and overviews on English derivational morphology or compounding. In such works, they are generally presented as a marginal topic that still falls under the scope of interest of overviews on word formation, but not as a prototypical case if we regard compounding and affixation / derivation as the main types of word formation processes.

In the literature on English word formation, the notion of combining forms was initially discussed only in relation to neoclassical compounds (as in Bauer 1983). In the past, linguists have sometimes claimed that final elements should not be described as combining forms (Quirk et al. 1985: 1520). The dominant view nowadays is that both initial and final combining forms display similar characteristics. Nevertheless, some final combining forms with a more general meaning are slightly more affix-like in terms of their semantics (cf. Fleischer & Barz 1995: 28 and Haspelmath 2002: 18 who regard a relatively abstract and general meaning as one of the typical features of affixes). We assume that from a diachronic perspective, there is a tendency for CFs to

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

bleach out in suffix position but to retain a more lexical status if they are used at the beginning of words.

Neoclassical compounding is a subfield of morphology that is also discussed in contrastive studies and literature on similar phenomena in other languages than (QJOLVK HJ /•GHOLQJHW DO RQ QHRFO Meesters (2004) on Dutch, Amiot & Dal (2007) on French, Iacobini (1997, 2010) on Italian, Petropoulou (2009) on neoclassical compounds in English and Modern Greek, DQG 3DQRFRYi RQ D FRQWUDVWLYH DQDO\VLV and Russian medical terminology. The more closely two European languages are 'genetically' related the higher the proportion of cognates they share. In Italian and other Romance languages, as well as in Modern Greek, the distinction between classic and modern word formation patterns is more blurred than, for instance, in Germanic or Slavic languages. Additionally, there is an effect of linguistic 'areality' within the group of European languages, i.e., the areal concentration of linguistic features in languages that found themselves in intensive contact situations (like English with large parts of its vocabulary derived from Germanic, Romance, and Graeco-Latin sources has a slightly special status among the Germanic languages. Classical and neoclassical elements (in many cases borrowed via French) are continuously in the process of merging with the vernacular.

It has been claimed that neoclassical formations in European languages have a number of peculiarities. There is some discussion on whether CFs, due to their similarity with affixes and with parts of regular compounds, fall under derivational or compounding processes. Neoclassical formations in particular do not clearly fall under one or the other major word formation processes, but the dominant view is that they are best treated as compounds and not as derivatives or cases of affixation (Bauer 2013: 441; 455f; Plag 2003: 74). The difficulty with this term also arises from a typological heterogeneity of the English word formation system, which allows both words and

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

stems as input to word formation processes. Combining forms have also sometimes been discussed under the topic of minor word formation types in English due to their potentially unclear status between elements in derivational or compounding patterns (e.g., Bauer 2006) or in the context of research on problematic or previously under researched areas in word formation theory (Elsen 2013a, 2013b). Other contexts in which English CFs are discussed are foreign and hybrid word formation (Eins 2008, 2009; /LQGQHU 5DLQHU 3|FNO Grammatical and Marginal Morphology morphology in connection with related phenomena such as blending (Fradin 2000; Mattiello 2013). Marginal morphology involves the discussion of grammatical, but non-prototypical processes, while extragrammatical morphology is associated with non-regular processes in artistic or playful use of language (Dressler 2000; Mattiello 2013: 28ff). Combining forms are not frequently given a section in their own right. Miller (2014: 207-219) devotes a chapter on formative extractions, combining forms, and neoclassical compounding. In the Oxford handbook of derivational morphology, there is a section summarizing bound roots, unique morphs, and neoclassical combining forms as different problematic morpheme types (Olsen 2015: 334-336). In the Oxford handbook of compounding (2009), combining forms are also briefly discussed in the section by Lieber (2009) as a problematic type of morphemes in complex words whose components do not occur outside the respective compounds; and neoclassical compounds are presented as a rather marginal class in the classification of compounds in the section by Scalise & Bisetto (2009). Despite being frequently discussed as marginal phenomena, according to Lieber (2009: 364), neoclassical compounds with combining forms continue to be coined productively in technical and medical fields. In several lexical fields numerous European languages share huge parts of their stock of neoclassical formatives due to a complex interplay of contact phenomena and common roots (Booij 2005: 87). Kastovsky (2009b: 326) suggests that this type of formations is on the increase in English as well as in all other European languages, particularly in technical jargon. Apart from such rough estimates of the productivity or frequency of

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

neoclassical formatives, many linguistic publications that address CFs remain rather theoretical in nature.

Recently, however, several academic publications on ongoing language change and English word formation in present-day English focus particularly on the productivity of novel native combining forms (e.g., Wiemer forthcoming), a subgroup of (mainly final) CFs, for which again different terms focussing on different aspects have been suggested, e.g. lexical affixes (Olsen 2014), unconventional suffixes or -fold morphs (Baldi & Dawar 2000), splinter-originating affixes (Danks 2003: 200ff), pseudo-suffixes (Kolin 1979), semi-suffixes or just suffixes. If they occur cross-linguistically as 'Euro-Anglicisms', they have been labelled, together with other internationalisms and 'Euro-Latin' / 'Euro-Greek' morphemes, as 'intermorphemes' (Kirkness 2005; Worbs 1995). Ultimately, of course, and from a larger perspective in the context of the group of the Indo-European languages, it can be argued that there is no sharp distinction between 'native' and 'non-native' elements if they have common ancestors and roots or have been borrowed from within the same language family. So-called native combining forms in English (cf. also Section 2 for examples) are a category of morphemes involved in creative word formation in a range of processes where shortening and compounding co-occur within lexemes. Lexemes with such formatives have often not yet been conventionalized. Nevertheless, they seem to behave similarly in certain aspects to the traditional CFs in neoclassical words and they all are efficient means of integrating lexical information into compact forms (Busse & Schneider 2007: 162).

Word formation involving native combining forms particularly shades off into blending. Blending involves the shortening of existing lexemes and may be accompanied by conceptual blending on the semantic level as well as the reduction of conceptual complexity by compression. Blends have been found to be a major source of new combining forms (Cannon 1986: 362). Therefore, native combining forms are most frequently discussed together with blending in current neologisms and

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

occasionalisms (e.g. Tomaszewicz 2008) and "the latest trends in English word formation" (Szymanek 2005). Danks (2003) suggested a detailed classification scheme to separate blending from other word formation processes, including compounding, neoclassical compounding, affixation, clipping and related phenomena. Blends are creative and relatively unpredictable formations that are coined through abstraction and comparison processes. From a structural point of view, such words are composed of two or more forms of which at least one has been reduced. Components of blends can be full words, initial splinters (splinters retaining the final part of the base), terminal splinters or mid splinters (splinters retaining the beginning or the middle of the base), splinter-originating affixes, shortened or entire initial or final combining forms, suffixes or phonesthemes (Danks 2003: 320). Phonesthemes, a term popularized by Firth (1930), are letter clusters below the morphological level, which evoke similar words with a similar meaning. Blends have different types of final components that may have been listed as combining forms in dictionaries such as the OED, but are described as different types of formatives in the literature.

Neoclassical morphemes are sometimes entirely excluded from recent studies on combining forms that are interested in current neologisms and occasionalisms of a certain type (e.g., Lehrer 1998; Mattiello 2017). The growing interest in new trends in word formation processes is the reason why the concept of combining forms has been given a very broad meaning with reference to both traditional combining forms, i.e. neoclassical word formation elements that are bound, and parts of words combined in recent blends. Recent 'native combining forms' can have a Germanic origin (-gate, -burger), but also a non-Germanic origin, e.g. shop<u>aholic</u>, talk<u>athon</u> or, Bowl<u>orama</u>, derived in analogy with more frequently used words such as alcoholic, marathon and panorama. They show a similarity with regard to the phonological and syllabic structure of their source words, but they can be freely combined with any element, preferably with native elements and entire words (shop, talk, bowl etc.). They undergo morphological reanalysis and may be subject to fashion or regional

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

preferences (e.g., coinages in American English hasheteria, candyteria, in analogy with cafeteria, evoking an exotic flair by adding a foreign ending to an English word). While a few older structures with native combining forms as bound elements (or sometimes simply referred to as suffixes if their lexical meaning has faded) such as otherwise or nowise are fossilized in standard English, new coinages such as educationwise with -wise in the sense of 'in the manner of', 'as regards' seem to be more productive in American English than in other varieties and strike some language critics as jargonistic. Such words may contain a syllable-forming linking vowel (e.g., in talkathon) but they do not have to if the initial part ends in a vowel (medievalism) or if they follow the syllabic structure and stress patterns of Germanic source words (e.g., snowscape in analogy with landscape, but cf. also waterscape, languagescape, cityscape as examples from the Corpus of Contemporary American English (Davies 2008-) with different syllabic structures).

It is outside the scope of this paper to discuss these forms in depth, but we deemed it necessary to make a few remarks on the subject of neo-combining forms as they have attracted increasing interest in the context of research on combining forms. We would like to exclude them from our analysis as they seem to be a recent phenomenon, and are not necessarily specifically related to word formation in scientific terminology and technical language as compared to the classic examples in which we are interested in our diachronic analysis of academic English. Native combining forms are primarily related to creative language and jargonistic expressions in media discourse, online media and advertisements. They may be formed by linguistic experimentation to be used in proper names, in brand names that succeed on international markets, in attention-catching book or film titles etc. or in spoken language where new words are coined all the time that do not fall under the most prominent word formation processes.

Various factors contribute to the fact that the category of combining forms may now appear slightly heterogeneous and that it includes elements of various origins that have

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

had different levels of productivity and technicality in different registers and ti periods. One of the reasons for this heterogeneity is that many linguistic publications on combining forms mainly address theoretical considerations. To date, not many experimental and corpus studies on combining forms have been published. This paper will have to narrow the concept down to prototypical forms of a very specific type to be used in a corpus analysis. We will focus on the morphological productivity of Graeco-Latin neoclassical combining forms from a diachronic perspective. These elements have played a major role for the coining and development of technical and scientific terms used in English academic writing. We will consider forms that typically fall under the topic of neoclassical compounding in the literature (Bauer et al. 2013: 441f), although it should be noted that neoclassical combining forms do not only occur in compounds, but in various formations in our data, e.g., in combination with other combining forms, with affixes or in combination with independent words. In various overviews and handbooks on morphology they may play a minor role, as these works do not particularly focus on languages for special purposes. However, for our dataset they play an essential role.

3. The role of combining forms in English scientific writing

This paper has its focus on the morphological productivity of combining forms in English scientific texts. Our case study in the following section has a particular focus on combining forms from the Graeco-Latin stock of lexical morphemes as we assume this to be a productive resource for one of the major word formation processes in English for specific purposes from the 17th century onwards. In many cases complex and productive word families of related lexemes based on such CFs have developed over time (cf. Busse 2002: 334). Some of these forms have already been used in English for a relatively long time, while new forms have been created more recently, e.g., the neoclassical element *cyber-*, which was formed within English in the 1960s as a combining form by clipping or shortening of the adjective *cybernetic* or the noun *cybernetics*, the science of automatic control systems. The same root can be found in

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

related to the Greek verb for 'to steer', also finds expression in English words such as governor or to govern that are less perceived as loans as they occur freely as independent lexemes and follow prototypical English inflectional and word formation paradigms. Similar sets of words typically appear in various languages and help to make the vocabulary used in English academic and technical writing more accessible to readers who may only have a smattering of the English language, but who can draw on their knowledge on cognates in inferring word meaning from other words that share these combining forms. Throughout the history of scientific innovation and discovery, the evolution of the language of science has given rise to a constant demand for new, and often morphologically complex words. In English, the development of former and current scientific terms is in many cases closely tied to morphological resources borrowed from Greek and Latin. Combining forms in neoclassical technical terminology seem to have always played a particularly important and productive role among English lexeme-formation elements, particularly for the creation of new nouns that facilitate the international communication of scholars.

It is not always possible to find out exactly how scientific terms were coined and by whom they were introduced, as there are often various persons who wished to claim ownership of an invention or are believed to have been the first to use a certain term. It is, for instance, sometimes reported that John Herschel was the first to use the word 'photography.' He made the term popular in English by using it in a paper read to the Royal Society of London in 1839, but it may well have been already in use in English before that date with the same sense or at least in the sense 'relating to the study of light'. It is possible that the German and / or French use of 'Photographie' / 'photographie' and of the corresponding adjectives predates the use of the terms in (QJOLVK 6RPH VRXUFHV VXJJHVW WKDW WKH DVWURQRPHU +HUVFKHO)RUHQFH KDG XVHG WKH WHUP in the 1840s, more than twenty people of different nationalities claimed to have invented photography (Warner Marien 2006: 15). In any case, the word started to replace various

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

competing terms derived from proper names, neoclassical or vernacular elements (e.g., photogenic drawing, heliograph, sun writing, sunpicture, daguerreotype, calotype, talbotype, F I (Q F \dFaCBFRtaSnnica 1859545).

In periods of stronger language purism in English and rejection of 'inkhorn terms' that were introduced by scholars in academic jargon, erudite coinages received some criticism as being pretentious, artificial or obscure for many readers, and in various cases alternative vernacular terms or literal English translations have been suggested to be used in academic texts or other registers. There is, for instance, the vernacular term loosestrife for a wildflower that coexists with the botanist term lysimachia. This plant was probably named after a king in ancient Greece, but the misinterpretation of the structure as a complex lexeme derived from λυ-, combining form of λύειν 'to loose', and μάχη 'strife' led to this English loan translation.

Another example for a literal English translation of a neoclassical term is the occasionally used pikeperch for a fish, nowadays probably better known as zander. It seems to be an English loan translation as a variant of the neo-Latin lucioperca (from Latin lucius 'pike' and perca 'perch'). It occurs in the Royal Society Corpus for instance as Lucioperca sandra, and in later corpus texts from other corpora such as the Corpus of Contemporary American English (Davies 2008) we sometimes find pikeperch (as a single word, as two words, or hyphenated). It is not always straightforward to recognize CFs or to distinguish between similar sounding CFs without detailed etymological knowledge, e.g., in lexemes with pedo- 'relating to soil', paedo-/pedo- in the sense of 'children' or pedo- in the sense of 'relating to the foot', cf. for instance orthopaedic/orthopedic – originally relating to the treatment of physical deformities, especially in children. The element -paed-, -ped- is frequently interpreted as deriving from the classical Latin word for 'foot' and has resulted in terms designating 'orthopaedic' footwear.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Early complex words with combining forms usually follow regular word formation rules and a conservative tradition for word formation patterns. At a time when the first academic publications were written in English, new terms were typically introduced by classically educated scholars. Many of the early formations have model words in German or French or are adaptations of words from scientific Latin. It has been claimed that from the Renaissance towards the early or mid-20th century, there was relatively little seepage of technical jargon into the English language at large, in contrast to later periods that were marked by a spread of technical and quasi-technical jargon into the media and everyday language (cf. Concise Oxford Companion to the English language 2005: 368). In early academic texts in English, combining forms of Greek origin were preferentially combined with other Greek elements (e.g., *biography*), and Latin combining forms with Latin elements (e.g., *agriculture*). Early scientific texts contain a relatively low number of hybrid forms. Newer coinages for technical terms, such as television were more freely coined in the form of hybrid lexemes composed of Greek and Latinate elements (e.g., if the 'pure' form, in this case *telescope*, had already been adopted with a different sense, cf. ibid.: 370), but the etymological distinction is not always clear-cut if cognates or early borrowings of Greek elements existed in classical and post-classical Latin or entered English via other European languages, e.g., via French such as *automobile* from Greek *auto-* 'self' and Latin *mobilis* 'movable' that was coined in France in the expression *voiture automobile* and subsequently borrowed into various other languages at the end of the 19th century.

In general, the productive use of combining forms in word formation processes is a means of avoiding alternative longer phrasal structures. This might be part of the explanation why phrasal scientific terms that were introduced in scientific English in the past sometimes tended to become later replaced by morphologically complex, but nevertheless compact words characterized by a potentially high semantic content per morpheme through the use of combining forms. This phenomenon can be observed in our corpus data for various structures (e.g., the phrasal structure *dephlogisticated air*

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

was replaced by scientists by the more compact word 'oxygen' (CF oxy- + -gen) with an intermediate stage where both forms were still in use). English terms consisting of combining forms contribute to linguistic economy and may serve as a base for further word formation processes such as derivation and inflection. For instance, we find the following set of word forms in our corpus data: electrophoresis, electrophoretic, electrophoretically, electrophoresed, and dielectrophoretic. Due to the dense encoding of information in such linguistic forms, there may be a certain potential of intransparency and ambiguity for readers who may also encounter terms such as electrophorus or electrophore in scientific texts (denoting a historical instrument for generating static electricity by induction, being only vaguely semantically related to the set of words that were given above in this paragraph and that are mainly used in more modern texts from the field of biology in our data). The potential ambiguity of such words is counterbalanced by their conventionalization in the language of science, the establishment of fixed terminology and more frequent use of formulaic structures over time. Nevertheless, not only in older scientific texts but also in recent specialized communication a certain degree of terminological variation can be observed. In some cases, CFs in term variants can occasionally be found in a different order (e.g. cardiovascular vs. vasculocardiac, cf. Bowker & Hakins 2006 who used the Web as a corpus as well as a medical database resource for specialized language investigations). Other terms may involve variants of the initial CF (e.g., orchiectomy, orchidectomy, orchiectomy, or testectomy, denoting the same type of surgical procedure) or of the final CF (e.g., achromatopsia, achromatopsy, and achromatopia having variants of the final elements). Additionally several longer term variants exist, such as achromatic vision or total colour blindness, with hybrid or Germanic origins, which can be used as medical and lay terms with the same meaning).

In general, precision of meaning and a high degree of information density in new terms are two potentially conflicting aspects motivating the coinage of words based on combining forms, but particularly in corpus texts from more recent periods,

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

compactness and vividness of expression are also relevant aspects that contribute to the prevalence of combining forms in technical English. The English language still uses many compounds and other types of complex words involving solely Greco-Latin elements, but certain hybrid forms with a combination of neoclassical combining forms and native English lexemes, terms or affixes have been coined as well. Additionally, particularly in recent texts from our dataset, we can observe some word formations with truncated forms of CFs in clippings and backformations resulting in words of different word classes and borderline cases between blends and compounds.

Lexemes with several combining forms are characterized by their high information density and an efficient combination of lexical morphemes within one orthographic word. In scientific writing, such lexemes are mostly technical terms. From a structural point of view, such compact technical terms avoid the use of alternative longer phrasal structures or multi-word terms. Some have suggested that technical terms in general are comparable to maximally condensed VHQWHQFHV RU WH[WV 3|FN not only encode the literal meaning of their components but also a conventionalized understanding of scientific phenomena. When a new technical term is introduced into the scientific community and becomes established, it is usually accompanied by a metacognitive discussion and a definition process that involves a conventionalization process. The use of terms in informationally dense texts works efficiently in scientific communication if the interlocutors can activate memorized contexts and if the terminology is consistent and transparent to the community. A higher transparency of technical terms can be facilitated by the use of certain word formation processes such as the use of combining forms. If they are familiar to the reader, their meaning is guessable from the literal meaning of their elements for those with a certain morphological and etymological awareness or knowledge of classical languages.

There are different degrees of perceived semantic transparency for individual interlocutors that we will not be able to address in this paper. For instance, in medical

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

terminology the frequent use of CFs facilitates the international communication of experts, but it can result in less transparent terms for lay audiences. The use of Graeco Latinate elements for describing pathological conditions or the anatomy of the human body also seems to have some stylistic functions as such elements may seem more elegant, polite or even euphemistic than elements of common language. Additionally, non-vernacular terms and expressions can influence the perception of lay people – a switch in terminology from medicalized terms to synonymous lay terms can result in a biased perception, for instance the assumption of how severe or how rare a medical disorder is, which has implications for medical communication with the public (cf. Young et al. 2008). In the future, we would like to complement our analysis with further work to find out whether technical words with CFs are holistically stored lexemes, how much information from word parts individuals integrate into their interpretation of a text and to what extent they recognize similarities between compound families with similar components and between scientific and lay terms.

It is not unusual for neoclassical compounds consisting of two CFs to become reduced to one morpheme consisting of one or two syllables. This reduction can be the result of the frequency of such compounds as lexemes in general language or in specialized technical language. The clipped form can be more colloquial on the one hand, but on the other hand it can also be a more technical version in a specific jargon. Both the clipped and the long form either continue to coexist in English or one starts to replace the other. The above-mentioned word graph representing chromograph is a result of fore-clipping as in phone from telephone, in which the beginning of the word has been dropped. Photo and bio are examples of back-clippings from photography and biography. Clipped neoclassical compounds typically have the same form as one of their component CFs, but they retain the meaning of a more complex word and are an efficient means of compressing lexical information into a highly compact unit. If both the combining form and a clipped version of a more complex lexeme with the same combining form coexist, the combining form itself is sometimes perceived to be on the

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

borderline between a free and a bound morpheme. The forms can potentially become semantically ambiguous or polysemous. Bio- can simply mean 'life' in the literal sense, but it also represents a reduced form of biographical (as in blends such as bio-sketch, i.e. a biographical sketch, or in blends such as bio-pic where both biographical and picture have been shortened) or of biological(ly) (e.g., in bio-degradable). Hydro-, for instance, can stand for 'water' or for hydroelectric (e.g., in hydropower) and auto- can represent the post-classical Latin and Greek word for 'self' or a shortened form of automobile.

As combining forms are on the borderline between affixes and words, we expect that CFs in our data can undergo reanalysis in grammaticalization and lexicalization processes and change their grammatical properties over time. In contrast to affixes, combining forms based on classical nouns can theoretically occur both in initial or final position. Some forms can be expected to occur productively only in one of these positions or change their productivity patterns and most frequent position in complex words over time. Like other types of lexical items they can become semantically bleached in diachronic data towards a more abstract and less lexical usage (e.g., final combining forms undergoing 'suffixization' and becoming markers of abstraction in lexical items). On the other hand, they can also move towards the other end of the scale and start to become more productive in the initial position of words, which – similarly to what can be observed with prefixes – is rather more lexical in nature than the final position, where suffixes and final elements generally tend to be more grammatical. A further grammatical shift in the use of combining forms towards a new grammatical function can lead to the creation of a separate lexical item.

Our case study in the next section will be illustrated by the combining form -lysis and related forms. We will start with a rationale and motivation of our study and the hypotheses that will drive our analyses (4.1). The methodological approach taken is described in Section 4.2. The results are presented in Section 4.3.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

4. Case study of combining forms in English scientific writing across time

4.1 Background

We are presenting the first stage of a larger experimental study, which covers various types of lexical and derivational morphemes in scientific English within a larger project on Information Density and Scientific Literacy in English (cf. Degaetano-Ortlieb & Teich 2016; Degaetano-Ortlieb et al. forthcoming).

In particular, we address word-internal complexity in a specific type of lexemes that consist of more than one morpheme of which at least one is a combining form. We include both initial and final root morphemes that mainly occur as bound lexical elements in English complex words, and we focus on those that are derived from nouns and verbs in classical languages, but not from adjectives (e.g., micro), prepositions or adverbs (e.g., hyper) or that go back to affixes in these languages. We also include truncated forms of such morphemes that can be attached to native or Latinate or Greek stems, to independent words of various origins, to affixes or to other truncated morphemes. The focus of this paper will be on illustrating the feasibility of our methods with some examples of complex lexemes, involving the neoclassical combining form -lysis, which can be WUDFHG EDFN WR *UHHN URRWV IU loosen' or *1 " ORRWVHQLQJ. We extend our analysis also to variants and related forms, such as reduced forms of these morphemes and combinations with additional affixes and linking elements (-lyt-, -lyz-, -lyst(s), -lytic(al) etc.). The list we use is based on the letter sequences extracted from lexemes in the OED WKDW KDYH *UHHN *1 the transliterated spelling of these Greek words in their etymological notes and on the list of variants of these morphemes in Robsen's compilation of Latin and Greek combining elements (1991: 88). For better readability, we will avoid referring to all individual forms in the following study. When we refer only to -lysis, we also refer to its variants.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

This set of forms can be assumed not to figure among the most productive combining forms in scientific English, but to have a rather average degree of productivity and to be used in technical terms in various texts throughout the diachronic stages of our corpus, denoting some type of decomposition, dissolution, loosening, breaking down or disintegration. Early formatives will presumably have a rather literal and unspecific meaning of 'breaking something into components', but words such as 'analyse' are not necessarily semantically transparent anymore, and in technical terms, the morpheme has a specific meaning, e.g., biological or chemical decomposition. Their productivity in word formation processes might increase or decrease with ongoing language change (cf. also Bauer 2001 and Haspelmath 2002: 114ff for discussions of different possibilities of measuring morphological productivity). This type of change might be accompanied by a change in perception so that the status of these morphemes and how native speakers perceive them is changing.

It is possible that some neoclassical terms have duplicative elements that may seem tautological from a structural point of view if they nearly have the same meaning but have been borrowed from different languages, e.g., in solvolysis or in solvolytic reaction from Latin solvere and Greek -lysis that both have a meaning related to the act of loosening or dissolution. This particular example is quite rare in very large corpora such as Davies' Google Books Corpus (2011) and does not occur in our dataset, but it is interesting to note that, although its parts are synonymous if considered in isolation, in combination with each other they denote a specific process of decomposition of a substance by the action of a solvent.

Moderntexts from the very recent past can be expected to be characterized by lexemes that reflect recent trends in word formation processes, such as clipping and blending involving combining forms, backformations with word-class changes and the development of independent lexemes from bound CFs (e.g., -lysis as a noun and -lyse as a verb from -lysis). Hybrid forms with CFs or reduced neoclassical compounds that are

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

combined with native English elements (e.g. *LysoTracker*, an example from our dataset from the 2000s, a trademark for a lysosome tracker) are equally a rather recent phenomenon.

The neoclassical forms derived from ancient Greek words have Indo-European (PIE) roots and are remotely related to other English words in common usage and cognates in the Romance and Germanic families (where we also find certain connections between verbs that mean 'to loosen' and 'to solve'). These can all be traced back to the PIE stem **leu-* 'to loosen, divide, cut apart', e.g. the English word *solve* from Latin *solvere* or the English word *lose* from OE *losian* (Partridge 1966: 1830ff). In contrast to these neoclassical combining forms, more naturalized words with the same origin can occur freely as independent lexemes and in various inflectional and word formation paradigms. There are different degrees of naturalization of English lexemes involving the neoclassical Gk *-lysis*. The status of this morpheme, its productivity, semantics and integration into the English language has changed over time.

4.2 Rationale and hypotheses

It has been suggested that speakers optimize communication through various types of reduction and by using structural cues that facilitate the predictability of upcoming elements (Levy & Jaeger 2007). Thus, words low in information content are more likely to become phonologically or structurally reduced or even omitted from their surrounding structures, while words with high information content generally tend to be longer, with the aim of maintaining a relatively uniform distribution of information in a text. It is assumed that the need to organize the contents of texts as efficiently as possible has led to the continuous adaptation of language usage and a selection of specific linguistic strategies. Texts of more recent scientific articles typically contain a higher average information density than texts from earlier periods (cf. e.g. Ortlieb et al. submitted). At the same time, processes of conventionalization in language use have

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

an influence on the expectations of text users with regard to upcoming linguistic elements based on previous experience with similar texts. The effect of information density on choices between different types of phonological or syntactic alternatives (e.g., the preferences of I'll for contracted auxiliaries or the use of complement clauses with and without the complementizer that) has been addressed by various scholars, especially in psycholinguistic studies (Demberg & Keller 2008; Hale 2001; Levy 2008) and recently also in corpus-linguistic studies (e.g., Degaetano-Ortlieb & Teich 2017; Schulz et al. 2016; Zimmerer et al. 2017). However, morphological structures within complex lexemes with the potential to combine various lexical morphemes still remain to be studied.

In our case study we consider one particular CF (-lysis) for investigating its diachronic development in scientific English. For this, we formulate the following hypotheses:

- H1: Conventionalized use of combining forms over time.
- H2: Interaction between convention and productivity.

In H1, we assume that as CFs enter language use, they will be increasingly used in specific (lexical or grammatical) contexts and adopt a conventionalized form over time, especially in scientific writing, where CFs are used to coin technical terms/terminology.

In H2, we assume that as these forms become conventionalized, their retrievability improves, allowing for more innovative uses and availability of these forms to be used in different yet closely related grammatical contexts (cf. De Smet 2016). This hypothesis is tested by considering conventionalized vs. productive use of CFs considering also word classes changes over time.

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

4.3 Methodology

In this section, we will present our data (4.3.1) as well as the extraction and analysis techniques adopted to test our hypotheses (4.3.2 ±4.3.5).

4.3.1 Data

We use two diachronic English corpora of scientific texts from various disciplines from the middle of the 17th century to the beginning of the 21st century ± the Royal Society Corpus (RSC; Kermes et al. 2016) and the Scientific Text Corpus (SciTex; De Castilho, Ortlieb et al. 2013) ± for analysing the evolution of scientific discourse with regard to the role of the combining form-lysis. The annotated corpora have been released in XML format and can be queried with CQP (Evert 2005), for instance via the 6 D D U E U • F N H Q &4 3 Z H E L Q W H U I D F H

The RSC consists of the digitized texts of the Philosophical Transactions and the Proceedings of the Royal Society of London published between 1665 and 1869. The Philosophical Transactions is the first and longest-running English scientific journal. The earliest of these journals covered all branches of science of the time. The Proceedings of the Royal Society also have a long history and commenced publication in the 19th century as a general science journal. The RSC version used in our study (v3.4) comprises about 32.5 million tokens. We plan to add to the corpus more digitized texts of publications by the Royal Society that were published at the end of the 19th century and the beginning of the 20th century.

These RSC data are complemented by the SciTex Corpus, a corpus of more contemporary texts of scientific English covering the 1970/80s and the early 2000s. SciTex consists of English scientific journal articles from several scientific disciplines. The current version contains approximately 39.2 million tokens. Table 1 shows the periods covered, the number of tokens as well as the number of documents in each time period. Both corpora are tokenized, lemmatized, and part-of-speech tagged. In

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

addition, each corpus contains metadata, e.g. on the author(s), discipline/topic, and year of publication.

Table 1: Corpus details

corpus	period	coverage	tokens	documents
RSC	1650	1665–1699	2,589,536	1,326
	1700	1700–1749	3,433,838	1,702
	1750	1750–1799	6,759,764	1,831
	1800	1800–1849	10,699,270	2,778
	1850	1850–1869	11,676,281	2,176
SciTex	1950	1966–1989	18,998,645	3,028
	2000	2000–2007	20,201,053	2,111

4.3.2 Selection and identification of combining forms

A selection of combining forms has been queried in our data to identify lexemes automatically that involve *-lyso-* / *-lysis* or variants of these combining forms and any other additional element(s), be they other root morphemes, affixes, or independent words (e.g., photo+lysis, para+lytic+al, re+ana+lysis, dia+lys+er, hydro+geno+lysis). Note that in our long-term study, of which we show here only a case study based on *-lysis* and related final elements, we considered larger variety of CFs.

Longer combining forms (e.g. *-membranaceo-*, listed as a combining form in the OED) typically tend to be less productive but more specific (the example *membranaceo-* occurs in the RSC, but it is rare and is typically used in scientific Latin terms in biology, e.g., *membranaceo-quadrangulatus*). Thus, we opted not to use very long CFs that are not very productive throughout our data or that will primarily lead to Latin quotations within our English corpus texts. However, several longer CFs that play no significant

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

role in early data become slightly more productive as a part of neoclassical compounds later in our data (e.g., anthropo- was first used productively between the 1840s and 1860s). Shorter forms are typically more productive, but are semantically less specific and sometimes have multiples senses. Thus, we also decided to exclude some CFs that were too short, and hence potentially more ambiguous, with less than three letters that are rather affix-like and that lead to too many irrelevant hits in the query results (e.g., ab-, formed within English as a combining form by clipping or shortening of 'absolute' in some electrical and magnetic units). We also excluded irrelevant words with unrelated similar sequences of letters and refined our queries in various ways to optimize the balance between precision and recall (e.g., the combining form -log- in a query with any preceding and following letters would also lead to forms such as Bologna – an example taken from our long-term study). The remaining CFs can be queried in our corpus as particular sequences of letters that occur within longer words. These forms consist of at least three letters that do not have a high ambiguity rate compared to shorter strings of characters for prefixes or suffixes or for strings of letters that represent native elements in English words. In our case study, we focus on -lysis and related final elements as a relatively unambiguous combining form of medium length.

4.3.3 Querying -lysis in corpus data

As a querytool, we use the Corpus Query Processor (CQP). We started by extracting -lys- and -lyz- in our data with the following query:

```
[word=".*ly[zs].* "].
```

The query results present a list of lexical items which were manually inspected. After refinements based on this list, the query was changed to

```
[word=".*ly[zs]i.*|.*ly[zs]e.*|.*lyte|.*lytic.*|.*lyst.*"].
```

The extracted list can be seen as a network of terms that are interrelated by sharing a morpheme and similar internal structures. Queries for CFs are based on strings of letters, but in contrast to queries for prefixes and suffixes they are usually not highly ambiguous. Only a few irrelevant words had to be sorted out manually that contained

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

such a string of letters representing different morphemes. There were a handful of occurrences of the word *proselyte* with a different etymology, denoting someone who converts to Judaism, and some occurrences of the word *aplysia*, a kind of sponge that does not include the morpheme we are interested in here. Some query hits were the result of OCR errors that we encounter more frequently in older texts from the 17th and 18th century, e.g., anatomical terms ending in *-pophysis* such as *zygapophysis*, *diapophysis*, *apophysis*, *metapophysis*, *prezygapophysis* that were wrongly recognized by the OCR software as *-spoplysis* in which we misleadingly obtained the sequence *-lysis*. We also excluded wrongly recognized strings of words that involve the correct morpheme (most frequently misspellings of *analysis* such as *ainalysis*, which could skew our results). In total, we had to exclude around 1.5% of all words from our query results.

4.3.4 Surprisal to measure convention vs. productivity

To observe whether a CF is used in a rather conventionalized vs. productive way, we measure the number of bits transmitted by these forms, i.e. their surprisal, which is formalized as $surprisal(CF) = -\log_2 p(CF|stem)$, where CF is the combining form and $p(CF|stem)$ the probability of a CF to occur with a particular stem. In other words, we consider how probable (or surprising) a particular CF is, given its stem. Note that we use the term 'stem' in our following analysis as a broad, overarching label for elements that are connected with CFs within complex words, most typically other bound stem or root morphemes with lexical meaning, but also any other elements that can be attached to CFs, for instance prefixes. If a particular combination occurs together frequently, the probability of having a particular CF combined with a particular stem is high (surprisal is low). A high number of low surprisal usages suggests a more conventionalized use. In contrast, in a rare combination of stem and CF, the probability of co-occurrence is low (surprisal is high). This indicates a more innovative use. A high number of innovative uses points to a higher productivity of a CF.

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

In our case study, besides considering the stem only, we also considered two words preceding the stem. This accounts for how probable a CF is, given its stem plus its preceding two words and allows us to better account for contextual information preceding the CF, i.e. whether a CF occurs in a predictive or less predictive context. This is formalized as: $surprisal(CF) = -\log_2 p(CF|stem, w_{stem-1}, w_{stem-2})$, where w_{stem-1} is the word preceding the stem and w_{stem-2} two words to the left of the stem.

To obtain these probabilities, we split the CFs from the stems and use a dedicated script for the calculation. The obtained surprisal values are annotated back into our corpus. For our analysis, we extract the stem, the combining form (cf), the part of speech (pos) of the lexical item (stem+CF), the surprisal value of the CF (srp_cf), and the time period these forms occur in (see Figure 1). In addition, we create bins of high, middle and low surprisal values to better compare the results across time periods. This categorization is based on a division of all values into quartiles (srp_cf_bins, again Figure 1). Bins of low surprisal will point to conventionalization, bins of high surprisal to productivity.

stem	cf	period	pos	srp_cf	srp_cf_bins
para	lytic	1750	JJ	0.48	low
ana	lytic	1750	JJ	5.63	high
para	lytic	1750	JJ	0.51	low
para	lytic	1750	JJ	0.21	low
para	lytic	1750	JJ	0.48	low
ana	lytic	1750	JJ	5.59	high
ana	lytical	1750	JJ	3.27	high
ana	lytical	1750	JJ	2.1	middle

Figure 1: Extraction results

4.3.5 Analysis

In our analysis, we pursue the above-mentioned hypotheses of conventionalized use of CFs over time (H1), on the one hand, and the interdependency of convention and productivity of CFs on the other, also considering changes in word class (H2).

ISSN 2453-8035 DOI:10.1515/larr-2017-0016

4.3.5.1 H1: Conventionalized use of combining forms over time

A conventionalized use of a particular CF with its stem will result in a relatively fixed use of the CF with one or few particular stems. To observe whether this is the case for -lysis, we consider the surprisal values of the CF, obtained by calculating how probable a CF is given its stem (see Section 4.2.4). The lower the surprisal, the less productive and more conventionalized the use of the CF is.

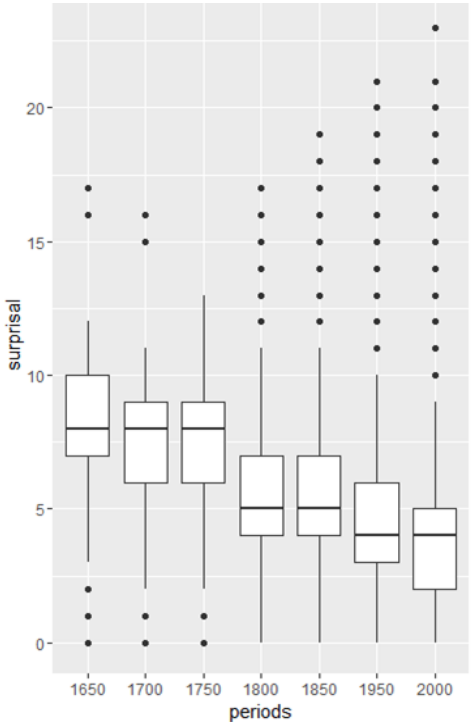


Figure 2: Surprisal distribution of -lysis across periods

Figure 2 shows the distribution of surprisal values of the -lysis CFs. Each boxplot gives information on the distribution for one particular time period. Outliers, which are observations distant from the general pattern of the distribution, are shown as black dots. The band inside the box represents the median. In Figure 2, between 1750 and 1800 and between 1850 and 1950, the distribution of surprisal values for -lysis changes. In the earlier time periods (1650–1750), surprisal is highest (with a median around 6). From 1800 onwards, surprisal drops significantly (with a median around 5). In the latest time periods (1950–2000), surprisal of -lysis forms achieves the lowest values (median of around 4). Thus, surprisal decreases significantly over time for the CF -lysis in scientific writing. This would confirm our hypothesis of a more conventionalized use

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

of this CF over the time periods inspected, i.e., in general, variation of the CF is reduced and its usage is confined to a use with particular preceding elements.

Besides this macro-analytic view, we can also inspect each instance at the micro-level, i.e. the particular forms that occur together. CFs with very low surprisal in the earlier periods are confined to analysis (1650-1750) as well as paralytic (1750). In the intermediate time periods (1800-1850), the noun analysis and the verb form analyses, analysed are used with very low surprisal as well as paralysis, catalytic, electrolyte and electrolysis. In the latest time periods, CFs with very low surprisal spread further to different word classes and forms (e.g., analysis, analyze, analyzed, analytically, analytic, analyses, analytical and analyzer) and to forms preceded by different elements (e.g., catalytic, dialysis, electrolyte, hydrolysis). Thus, while in the earlier time periods, analysis was the only very predictive combination, in the later periods -lysis combined with ana- is used in different word classes and becomes also predictable with other stems. This indicates that while there is a conventionalized use of different -lysis forms combined with only particular stems, the CF itself seems to pass through phases of productive use.

In addition, we see from Figure 2 more outliers at the higher ends towards the later periods (1800-2000). While the general pattern of change for the -lysis reflects a more conventionalized use confined to particular stems, some uses are of a more innovative kind. This means that the CF is hardly predictable on basis of the preceding element, indicated by a high surprisal value of the CF, i.e., stem and CF seldom appear together. These are forms such as non-analytical (the CF having a surprisal value of ~14, example 1), and terms such as histolytica (surprisal of ~23, example 2) and FE-analysis (surprisal of ~10, example 3).

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

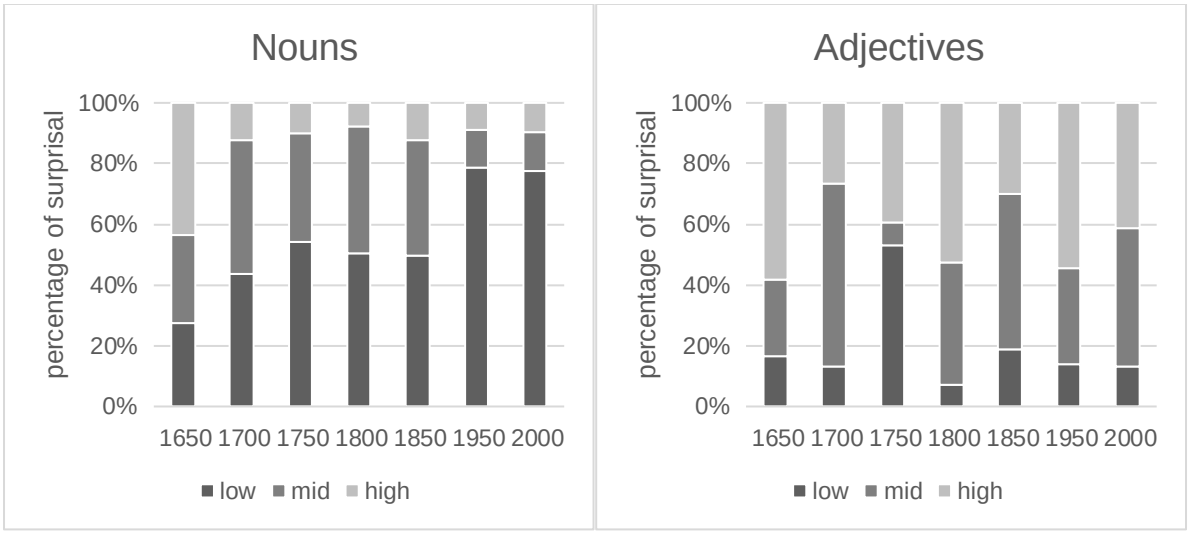
(1) Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to Entamoeba histolytica (2000, scientific discipline of biology)

(2) This requires a calculation of the objective function and FE-analysis to verify the constraints (2000, scientific discipline of digital construction)

(3) When approaching the discretization of the biharmonic equation with analytical procedures, there are mainly two options (2000, scientific discipline of digital construction)

4.3.5.2 H2: Interplay between convention and productivity

The earliest use of the CF -lysis in our dataset is the noun analysis with a general meaning of 'examination' or 'study' that may have sparked the introduction of more specific and technical terms with this morpheme. While use in nouns continues to persist over time, the CF becomes productive in other word classes as well: adjective, verb and adverb. To capture this development, we consider surprisal values of the CF based on its preceding context (stem + two previous words), comparing low, middle and high bins of surprisal across time and word class. Again, a high number of low surprisal values indicates a conventionalized use, while a high number of high surprisal values indicates productive use of the CF.



ISSN 2453-8035 DOI:10.1515/lart-2017-0016

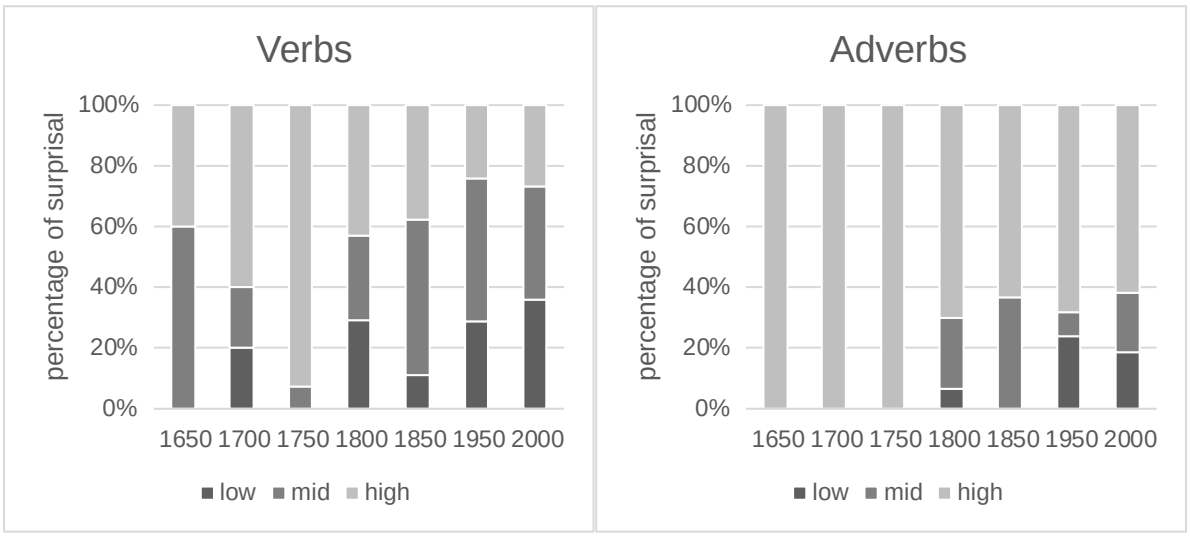


Figure 3: Percentage of surprisal in low, middle and high bins across time for different word classes of -lysis

From Figure 3, we can see how the percentage of low, middle and high surprisal varies across word classes and time periods. For nouns, we see an increase of low surprisal over time to almost 80%. This indicates a quite conventionalized use of the CF -lysis, mainly confined to analysis, which becomes more and more predictable over time. Nouns with high surprisal are analyser, analyst, paralysis (1650–1850), as well as plural forms such as analyses, analysts and paralytics. New forms arise in 1800–1850 with analyser, catalysis, electrolysis, electrolytes, and dialys/zer. From 1950, the productivity of the stem rises considerably, especially with the base form -lysis (e.g., acido-, hydrogeno-, nucleo-, radio-, methano-, psychoana-, thermolysis, etc.), with plural forms (e.g., photolyses, pyrolyses), with the ending -lysine (e.g., poly-, chromatin-poly-D-, DNA-poly-D-lysine), as well as with forms occurring in hyphenated compounds (e.g., chromosome-analysis, texture-analysis, free-text-analysis, flow-analysis, FE-analysis). Note that these relatively innovative hyphenated forms are combined with the very predictive noun analysis. This would confirm De Smet (2016)'s observation that as a form achieves a status of high conventional use, with a highly improved mental retrievability, more innovative uses can be generated.

Considering the surprisal of adjectives (see again Figure 3), low surprisal values are mostly below 20%, i.e. the use of -lysis variants in adjectives is relatively unpredictable

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

showing high variation with respect to its stems. Interestingly, for the adjective there is a rise of middle surprisal in 1700, which seems to move to an increase of low surprisal in 1750. Considering the lexical realizations, these effects are related to the adjective *paralytic*, which has high surprisal in 1650 (3.22), middle in 1700 (1.39), and low in 1750 (0.07). Thus, as *paralytic* moves into language use, it becomes more and more predictable. If we track its further development, *paralytic* is rarely used from 1800 onwards, reflected in its surprisal value going up again over time (1800: 2.56, 1850: 6.39, 1950: 14.27, 2000: 14.13). Thus, the use of the CF *-lysis* as an adjective remains relatively variable over time, with no clear conventionalized lexical realization.

With regard to verbs, Figure 3 shows an increasing tendency towards high surprisal values from 1650 to 1750 and then a decrease of high and an increase of low surprisal values for the later time periods. In the earlier periods, *analyse* is used in various verb forms (*analysing*, *analysed*, and *analys/ze*) and with different surprisal values, the rise in high surprisal towards 1750 being due to this kind of variation, pointing to a productive use. From 1800 onwards, the past tense form (*analysed*) has a low surprisal value, while the other forms show middle to high surprisal values. In addition, the CF is used with *electro-* in *electrolyse*, *electrolyzed*, *electrolys/zing* with high surprisal. In 1850, *analyse* in the past tense becomes relatively predictable showing low surprisal, and new forms in the past tense combined with different stems emerge (e.g., *catalysed* and *dialyzed*, with high surprisal). In 1950, *analys/ze(s)* and *catalys/ze(s)* in the present tense become quite predictable (low surprisal of approx. 0.36) as well as the past tense forms of *ana-*, *cata-* and *dia-* and *hydrolys/zed* showing low surprisal values (around 0.38). Among the forms with high surprisal we find *alkalilysed*, *phosphorolyzed*, *photolys/zed*, as well as the *-ing* form *analys/zing*. In 2000, *analyse* in various verb forms is quite predictable (low surprisal), while most forms with high surprisal are *co-analyzed*, *preanalyze*, *proteolysed*, *re-analyzed* and *-ing* forms such as *catalysing* and *paralyzing*. Again, we see how forms establish themselves into language use while

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

new forms arise. Most of these new forms are either new verb forms such as -ing forms or conventionalized verb forms such as past tense forms derived from analysed, which then have an increase in productivity with new stems.

As for adverbs (see again Fig. 3), we clearly see how unpredictable they are in earlier time periods in the data. Only from 1800 onwards, adverbial forms of variants of the CF -lysis enter language use in a somewhat conventionalized way. From 1650 to 1750, analytically is the only lexical realization. In 1800, analytically moves also to middle and low surprisal uses, and the form electrolytically arises. In 1850, catalytically enters language use. In 1950, analytically becomes quite predictable and new forms arise (cyto-, endonuclease, exonuclease, hydro-, proteolytically). In 2000, some of these forms move to middle surprisal (e.g. proteolytically), while new forms again enter language use (e.g. autocatalytically).

In summary, while the original base form analysis of the CF -lysis becomes more and more predictable over time with an established meaning and use, new forms such as the plural form or nouns such as analyser, arise over time. Moreover, as analysis becomes more conventionalized, it spreads out to other word classes. The use within these word classes shows a similar tendency: as some forms become established, new forms arise within a word class.

5. Conclusions and outlook

In this paper, we have investigated the diachronic development of a CF (-lysis) and its variants in English scientific texts over a period of approximately 350 years. CFs are used as a word formation process for expressing information in a condensed way, and are therefore particularly useful in scientific texts. After setting the scene on the category of CFs, their status and their role in English scientific writing, we have presented a case study on the CF -lysis. The study is part of a larger project in which we assume that English scientific writing has become more informationally dense over

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

time. CFs are one of many possible phenomena that facilitate a more informationally dense linguistic encoding. In particular, we were interested in the morphological productivity of *-lysis* and its diachronic course of change. In our analysis, we compared a conventionalized vs. a productive usage of CFs over time (Section 4.3–5.1) and the interaction between convention and productivity across the word classes carrying lexical meaning (Section 4–5.2).

To measure a more conventionalized vs. a more productive use, we use the notion of surprisal to measure how probable a particular CF is given its stem. The higher the number of probable combinations, the more conventionalized the CF is, while the higher the number of less probable combinations, the more productive the CF is. Surprisal has the advantage of accounting for probabilities conditioned on a context (here previous context) which cannot be achieved by considering mere frequencies (i.e. unconditioned probabilities).

Firstly we have observed that the use of *-lysis* becomes more conventionalized over time, i.e. particular forms of *-lysis*, especially noun forms, are increasingly used with the same stems. Secondly, since *-lysis* as a noun becomes more conventionalized, it is increasingly used in other related word classes: first adjectives, followed by verbs and then adverbs. This result confirms the above-mentioned observation by De Smet (2016). Thirdly, the use within these word classes shows how some forms become established with a relatively conventionalized use allowing new forms to arise within a word class with an increase in productivity.

In our further research, we plan to extend this case study with an analysis of a larger set of combining forms to generalize our findings. We would also like to apply our approach to other modern or historical monolingual or multilingual corpus data across different registers and text types (e.g., the synchronic bilingual GECCo corpus with texts from a wide range of written and spoken registers and text types, cf. Metzler &

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Lapshinova-Koltunski 2014 and Menzel 2016). Additionally, we would like to produce corpus-based dictionaries of technical terms involving combining forms and then thematically or chronologically via automatic methods in combination with our corpus metadata.

Abbreviations

CF – combining form

CQP – Corpus Query Processor

EFL – English as a foreign language

OED – Oxford English Dictionary

RSC – Royal Society Corpus

SciTex – Scientific Text Corpus

References

A thesaurus of English word roots (2014). Danner, H.G. (ed.) Lanham: Rowman & Littlefield.

Amiot, D. & Dal, G. (2007). Integrating neoclassical combining forms into a lexeme-based morphology. In Proceedings of the 5th Mediterranean Morphology Meeting (MMM5),) U p M X 1 8 September 2005, p. 322-336.

Ayers, D.M. (1965). English words from Latin and Greek elements. Tucson: University of Arizona Press.

Ayers, D.M. (1972). Bioscientific terminology. Words from Latin and Greek stems. Tucson: University of Arizona Press.

Baldi, P. & Dawar, C. (2000). Creative processes. In Morphology. An international handbook on inflection and word formation, 1. Booij, G.E., Lehmann, C. & Mugdan, J. (eds.) Berlin: De Gruyter, p. 963-972.

Bauer, L. (1983). English word formation. Cambridge: Cambridge University Press.

Bauer, L. (1998). Is there a class of neoclassical compounds, and if so is it productive? In Linguistics, 36 (3), p. 403-422.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Bauer, L. (2001). Morphological productivity. Cambridge: Cambridge University Press.

Bauer, L. (2006). Compounds and minor word formation types. In: The handbook of English linguistics. Aarts, B. & McMahon A. (eds.). Malden, MA: Blackwell. p. 483–506.

Bauer, L., Lieber, R. & Plag, I. (2013). Oxford reference guide to English morphology. Oxford: Oxford University Press.

Booij, G. (2005). The grammar of words. An introduction to linguistic morphology. Oxford: Oxford University Press.

Bowker, L. & Hawkins, S. (2006). Variation in the organization of medical terms ± Exploring some motivations for term choice. Terminology 12 (1), p. 79–110.

Busse, U. (2002). Case study III: English, In: Lexicology. An international handbook on the nature and structure of words and vocabularies. [HSK 21, 1]. Cruse, D.A., Hundsnurscher, F., Job, M. et al. (eds.). Berlin-New York: de Gruyter, p. 828–836.

% X V V H 8   6 F K Q H L G H U '   . • U] H Sprachliche J O L V *Kürze: Konzeptuelle, strukturelle und pragmatische Aspekte*. (= Linguistik ± Impulse & Tendenzen, 27). % l U - $ 5 R H O F N H 7   6 W H L Q K D X H U $ Berlin-New York: De Gruyter, p. 159–180.

Cannon, G. (1986). Blends in English word formation. In: Linguistics, 24, p. 725–753.

Concise Oxford companion to the English language. (2005). McArthur, T. (ed.) Oxford: Oxford University Press.

Danks, D. (2003). Separating blends: a formal investigation of the blending process in English and its relationship to associated word formation processes. PhD thesis, University of Liverpool. Available at: http://rdues.bcu.ac.uk/debbiedanks.html

Davies, M. (2008) The corpus of contemporary American English (COCA): 520 million words, 1990–present. Available at: http://corpus.byu.edu/coca/

Davies, M. (2011) The Google Books Corpus. Based on Google Books n-Grams. Available at: http://googlebooks.byu.edu/

 ISSN 2453-8035 DOI: 10.1515/lart-2017-0016

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. In Language variation and change, 28 (1), p. 83-102.

Degaetano-Ortlieb, S., Kermes, H., Lapshinova-Koltunski, E. et al. (2013). SciTeX - A diachronic corpus for analyzing the development of scientific registers. In New methods in historical corpus linguistics. Corpus linguistics and interdisciplinary perspectives on language - CLIP, 3. Bennett, P., Durrell, M., Scheible, S. et al. (eds.). 7 • E L Q J H Q 1 1 0 4 U S

Degaetano-Ortlieb, S. & Teich, E. (2016). Information-based modelling of diachronic linguistic change: from typicality to productivity. In Proceedings of Language technologies for the socio-economic sciences and humanities (LATECH'16), Association for Computational Linguistics (ACL), 7-12 August 2016, Berlin, Germany, p. 165-173.

Degaetano-Ortlieb, S. & Teich, E. (2017). Modeling intra-textual variation with entropy and surprisal: topical vs. stylistic patterns. In Proceedings of Language technologies for the socio-economic sciences and humanities (LATECH'17), Association for Computational Linguistics (ACL), Vancouver, Canada, 30 July - 4 August 2017, p. 68-77.

Degaetano-Ortlieb, S., Kermes, H., Khamis, A. et al. (forthcoming). An information-theoretic approach to modeling diachronic change in scientific English. In Selected papers from Varieng - From data to evidence (d2e), Brill, Language and Computers.

Degaetano-Ortlieb, S., Kermes, H., Teich, E. (submitted). Capturing diachronic register variation using entropy and surprisal. In Digital scholarship in the humanities (DSH). Oxford Academic.

Demberg, V. & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. In Cognition, 109, p. 193-210.

Denning, K., Kessler, B. & Leben, W.R. (2007). English vocabulary elements, 2nd ed. Oxford-New York: Oxford University Press.

Dictionary of English word-roots: English-roots and roots English, with examples and exercises (1969). Smith R.W.L. Totowa: Littlefield.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Dictionary of word roots and combining forms, compiled from the Greek, Latin, and other languages, with special reference to biological terms and scientific names. (1960). Borror, D.J. (ed.). Mountain View, CA: Mayfield Publishing Company.

Dofka, C.M. (2013). Dental terminology. 3rd ed. Clifton Park, N.Y.: Delmar: Cengage Learning.

Donalies, E. (2000). Das Konfix. Zur Definition einer zentralen Einheit der deutschen Wortbildung. In Deutsche Sprache, 28, p. 144-159.

'RQDOLHV (6WLHIOLFKHV *HRIDV] Studien DLQP zur Fremdwortbildung. (= Germanistische Linguistik 197/198), 0•OOHU 3 2 H Hildesheim: Olms, p. 41-64.

Dressler, W.U. (2000). Extragrammatical vs. marginal morphology. In Extragrammatical and marginal morphology. Doleschal, U. & Thornton, A.M. (eds.). Munich: Lincom, p. 1-10.

Durkin, P. (1999). Root and branch: revising the etymological component of the Oxford English Dictionary, In Transactions of the Philological Society, 97, p. 1-49.

Eins, W. (2008). Muster und Konstituenten der Lehnwortbildung. Das Konfix-Konzept und seine Grenzen. Hildesheim et al.: Olms.

(LQV : $OWHU :HLQ LQ QHXHQ Studien zu 6FKOlXFKH Fremdwortbildung. 0•OOHU 3 2 (ed.). Hildesheim et al.: Olms, p. 65-8.

Eins, W. (2015). Types of foreign word formation. In Word formation. An international handbook of the languages of Europe. 0•OOHU 3 2 2KQKHLVHU , 2 Berlin-Boston: de Gruyter, p. 1561-1579.

Eisenberg, P. (2012). Das Fremdwort im Deutschen. 2nd ed. Berlin: de Gruyter.

Elsen, H. (2005). Deutsche Konfixe. In Deutsche Sprache, 33 (2), p. 133-140.

Elsen, H. (2013a). Problemzonen der Wortbildung. In G HU (LQWUDJ LP :|U Wortbildung im elektronischen Wörterbuch. Schriften des Instituts für Deutsche Sprache. Klosa, A. (ed.). 7•ELQJHQ 1DUU S 1-103.

Elsen, H. (2013b). Zwischen Simplex und komplexem Wort ± eine holistische Sichtweise. In *"Wenn die Ränder ins Zentrum drängen". Außenseiter in der*

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Wortbildung(sforschung). Born - 3 | F N O H G V Berlin: Frank & Timme, p. 25-42.

Evert, S. (2005). The CQP query language tutorial. IMS: Stuttgart University.

Firth, J. R. (1930). Speech. London: Oxford University Press.

Fleischer, W. (1995). Konfixe. In *Wort und Wortschatz. Beiträge zur Lexikologie*. Pohl, I. & Ehrhardt, H. (eds.). 7 • E L Q J H Q 1 L H P H \ H U S. 61-68.

Fleischer, W. & Barz, I. (1995). Wortbildung der deutschen Gegenwartssprache. 7 • E L Q J H Q 1 L H P H \ H U

Fradin, B. (2000). Combining forms, blends and related phenomena. In Extragrammatical morphology and marginal morphology. Doleschal, U. & Thornton A.M. (eds.). Munich: Lincom, p. 11-59.

Grimm, H.-J. (1997). Konfixe. Beobachtungen in Tageszeitungen und in : | U W H U E X F K H U Q Nominationsforschung im Deutschen. % D U] , 6 F K U | G H (eds.). Frankfurt/M.: Peter Lang, p. 277-284.

Google Books Ngram Viewer. Available at: http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

Guide to the third edition of the OED. Available at: http://public.oed.com/the-oed-today/guide-to-the-third-edition-of-the-oed/

Hacken, P. ten (1994). Defining morphology: A principled approach to determining the boundaries of compounding, derivation, and inflection. Hildesheim: Olms.

+ D F N H Q 3 W H Q 3 D Q R E R Y c 1 Neoclassical formatives in dictionaries. In Proceedings of the 16th EURALEX International Congress, Bolzano, Italy, 15-19 July 2014, p. 1059-1072.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL '01), Pittsburgh, USA, 1-7 June 2001. Stroudsburg, PA: Association for Computational Linguistics, p. 1-8.

Hamans, C. (2014). The status of -o- or on the allomorphy of neoclassical compounds. In Linguistic insights: studies on languages. & U X] & D E D Q L O O D V , , 7 H

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

C. (eds.).$ O F D O i G H + H Q D U H V 8 Q L Y H U V L G D G G H $ O F D C
217.
Haspelmath, M. (2002). Understanding morphology. London: Arnold.
How to use the OED – Key to frequency. Available at:
http://public.oed.com/how-to-use-the-oed/key-to-frequency/
Iacobini, C. (1997). Distinguishing derivational prefixes from initial combining forms.
In Proceedings of 1st Mediterranean Morphology Meeting, Mytilene, Greece, 19-21
September 1997, p. 132-140.
Iacobini, C. & Giuliani, A. (2010). A multidimensional approach to the classification
of combining forms. In Italian journal of linguistics, 22 (2), p. 278-316.
Kastovsky, D. (2009a). 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-EX), Lammi, Finland, 25-28 April
2008, p. 1-13.
Kastovsky, D. (2009b). Diachronic perspectives. The Oxford handbook of
compounding / L H E H U 5 â W H N D Oxford H Oxford 3 Hniversity Press, p.
323-340.
Kean, D. 2017. 30 January. "Oxford dictionary considers including wave of Trumpian
neologisms" The Guardian. Available at: https://www.theguardian.com/books
/2017/jan/30/oxford-dictionary-donald-trump-neologisms
Kermes, H., Knappen, J., Degaetano-Ortlieb, S. et al (2016). The Royal Society
Corpus: From uncharted data to corpus. In Proceedings of the 10th International
Conference on Language Resources and Evaluation (LREC 2016) U W R U R å 6 O R
23-28 May 2016, p. 1928-1931.
Kirkness, A. (2005) Zur lexikographischen Dokumentation eurolateinischer
Wortbildungseinheiten. Vergleichende Beobachtungen am Beispiel -. atio
Fremdwortbildung. Theorie und Praxis in Geschichte und Gegenwart O O H U 3 2
(ed.). Frankfurt/M.: Lang, p. 447-483.
Kolin, P. C. (1979). The pseudo-suffix -oholic. In: American Speech, 54 (1), p. 74-76.

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

Lehrer, A. (1998). Scapes, holics and thons: The semantics of English combining forms. In: American Speech, 73(1), p. 3-28.

Levy, R. (2008). Expectation-based syntactic comprehension. In: Cognition, 106(3), p. 1126-1177.

Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In: Advances in neural information processing systems (NIPS), 6 F K O | N R S I % 3 O D W W + R I P D Q Q 7 H G V Cambridge, MA: MIT Press, p. 849-856.

Lieber, R. (2009). A lexical semantic approach to compounding. In: The Oxford handbook of compounding. Lieber, R. â W H N D X H U 3 + H G V Oxford: Oxford University Press, p. 434-452.

Lindner, T. & Rainer, F. (2015). Word formation in Neo-Latin. In: Word formation. An international handbook of the languages of Europe. 0• O O H U 3 2 2 K Q K H L V S. et al. (eds.). Berlin & Boston: de Gruyter, p. 1580-1597.

/ • G H O L Q J $ 6 F K P L G 7 . L N R . 5 S D V R J O R X 6 Neoclassical word formation in German. In: Yearbook of morphology. Booij, G., & Marle, J. van (eds.). Berlin: Springer Netherland, p. 253-283.

Maier, E.: The 'gate' suffix. Available at: http://public.oed.com/aspects-of-english/english-in-use/the-gate-suffix/

Marchand, H. (1969). The categories and types of present-day English word formation – A synchronic-diachronic approach. 2nd ed. Munich: C.H. Beck.

Mattiello, E. (2013). Extra-grammatical morphology in English. Abbreviations, blends, reduplicatives, and related phenomena. Berlin & Boston: de Gruyter

Mattiello, E. (2017). Analogy in word formation: A study of English neologisms and occasionalisms. Berlin & Boston: de Gruyter.

Macmillan Open Dictionary: 'combining form'. Available at: http://www.macmillandictionary.com/dictionary/british/combining-form

McCauley, J. (2006). Technical combining forms in the third edition of the OED: Word formation in a historical dictionary. In: Selected Proceedings of the 2005 Symposium

 ISSN 2453-8035 DOI: 10.1515/lart-2017-0016

on New Approaches in English Historical Lexis (HEL-LEX), Helsinki, Finland, 17-19 March 2005, p. 95-104.

Meesters, G. (2004). Marginale morfologie in het Nederlands. Paradigmatische samenstellingen, neoklassieke composita en splintercomposita. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

Menzel, K. & Lapshinova-Koltunski, E. (2014). Kontrastive Analyse deutscher und HQJOLVFKHU .RKlVLRQVPLWWHO LQ Tekst i dyskurs-Text und Diskurs. *Zeitschrift der Abteilung für germanistische Sprachwissenschaft des* Germanistischen Instituts Warschau, 7, p. 247-266.

Menzel, K. (2016). The elusive ellipsis - the complex history of a vague grammatical concept in need of empirical grounding. In Lege artis. Language yesterday, today, tomorrow, The journal of the University of SS Cyril and Methodius in Trnava, 1 (1), June 2016, p. 163-201. DOI: 10.1515/lart-2016-0004

Merriam-Webster Online. Springfield, MA: Merriam-Webster, 'combining form', Available at: https://www.merriam-webster.com/dictionary/combining%20form

0LFKHO 6 'DV .RQILI]ZLVFKHQ /DQJXH XQ sprachgebrauchsbezogenen Definition und Typologie. In Studien zur Fremdwortbildung. (= Germanistische Linguistik 197/198) 0ller, P.O. (ed.). Hildesheim et al.: Olms, p. 91-140.

Miller, G. (2014). English lexicogenesis. Oxford: Oxford University Press.

0•O-OHSLU W]HU & :|UWHUEXFKHU Internetwörterbücher. Ein Kompendium. .ORVD $ -Spitzer, C. (eds.) Berlin-Boston: de Gruyter, p. 291-342.

Mugdan, J. (1989). Grammar in dictionaries of languages for special purposes (LSP). In Hermes, 3, p. 125-142.

Nord, W.R. & Connell, A.F. (2011). Rethinking the knowledge controversy in organization studies: A generative uncertainty perspective. New York: Routledge.

Ologies and isms: A dictionary of word beginnings and endings. (2005). Quinion, M. (ed.). Oxford: Oxford University Press.

-Ologies and-Isms: A thematic dictionary (1986). 3rd ed. Urdang, L. (ed.), Detroit: Gale Research.

Olsen, S. (2014). Delineating derivation and compounding. In: The Oxford handbook of derivational morphology. Lieber, R. & Stekauer, P. (eds.). Oxford: Oxford University Press, p. 246-269.

Oxford English Dictionary. (1884-1928, 1st ed. ('A New English Dictionary on Historical Principles') / 1989 2nd ed., 2000- 3rd ed.), Oxford: Clarendon Press. 3rd ed. Available at: http://www.oed.com/

Oxford English Dictionary Online, 'Advanced Search'. Available at: http://www.oed.com/advancedsearch

Oxford English Dictionary Online, Advanced Search results for 'combining forms' Available at: http://www.oed.com/search?caseSensitive=true&nearDistance=1&ordered=false&pos_0=combining+form&scope=ENTRY

Oxford Dictionaries, 'combining form'. Available at: https://en.oxforddictionaries.com/definition/combining_form

Oxford English Dictionary Online, 'lysis', comb. form. Available at: http://www.oed.com/view/Entry/111701

Oxford English Dictionary Online, 'lyso-', comb. form. Available at: http://www.oed.com/view/Entry/111702

3 D Q R F R Y i 5 Categories of word formation and borrowing: an onomasiological account of neoclassical formations. Newcastle upon Tyne: Cambridge Scholars Publishing.

Partridge, E. (1966). Origins: A short etymological dictionary of Modern English. 4th ed. London & New York: Routledge.

Petropoulou, E. (2009): On the parallel between neoclassical compounds in English and Modern Greek. In: Patras Working Papers in Linguistics, 1, p. 40-58.

Plag I. (2003). Word formation in English. Cambridge: Cambridge University Press.

3 | F N O : Le confixe: element rebelle a vocation internationale. In: Actes du

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

XXVe Congrès de Linguistique et de Philologie Romanes, 7, Innsbruck, Austria, 3-8 September 2007. Berlin: De Gruyter, p. 471-476.

3 | F N O : Konfixe ± H L Q X Q H U V F K | S I O L F K H U 9 R U U D W Fachsprache(n) in der Romania – *Entwicklung, Verwendung, Übersetzung*. Sergo, L. & Wienen, U. (eds.) Berlin: Frank & Timme, p. 97-109.

3 | F N O W (2015). Foreign word formation, language planning and purism. In Word formation. An international handbook of the languages of Europe. 0 • O O H U 3 2 Ohnheiser, I., Olsen, S. et al. (eds.). Berlin & Boston: De Gruyter, p. 1598-1614.

3 U ü L ü 7 he treatment of affixes in the 'Big Four' EFL dictionaries. In International journal of lexicography, 12, p. 263-279.

3 U ü L ü 7 + H D G K R R G R I V X I I L [H V D Q G I L Q D O formation. In Acta linguistica Hungarica, 54, p. 381-392.

3 U ü L ü 7 3 U H I L [H V Y V L Q L W L D O lexicographic F R P E L Q L Q J perspective. In International journal of lexicography, 18, p. 313-334.

3 U ü L ü 7 6 X I I L [H V Y V I L Q D O F R P E L Q L Q J I R perspective. In International journal of lexicography, 21(1), p. 1-22.

Prefixes and other word-initial elements of English (1984). Urdang, L. (ed.). Detroit: Gale Research.

Quirk, R., Greenbaum, S., Svartvik, J. et al. (1985). A comprehensive grammar of the English language. London: Longman.

5 D G L P V N ê Noun+noun compounds in Italian: A corpus-based study. ý H V N p % X G M R Y L F H - L K R þ H V N i X Q L Y H U] L W D

Robertson's words for a modern age. A cross reference of Latin and Greek combining elements (1991). Robertson, J.G. (ed.). Eugene, OR: Senior Scribe Publications.

Saarbrücken CQPweb. Available at: http://corpora.clarin-d.uni-saarland.de/cqpweb/

Scalise, S. & Bisetto, A. (2009). The classification of compounds. In The Oxford handbook of compounding. / L H E H U 5 Pâ (eds.) 2 [I R U G Oxford: Oxford University Press, p. 34-53.

 ISSN 2453-8035 DOI: 10.1515/lart-2017-0016

Schulz, E., Oh, Y.M., Malisz, Z. et al. (2016). Impact of prosodic structure and information density on vowel space size. In Proceedings of speech prosody 2016. Boston, USA, 31 May-3 June 2016, p. 350-354.

Seiffert, A. (2008). Autonomie und Isonomie fremder und indigener Wortbildung am *Beispiel ausgewählter numerativer Wortbildungseinheiten*. Berlin: Frank & Timme.

Shannon, C.E. (1948). A mathematical theory of communication. In Bell syst tech journal, 27, p. 379-423 (Part I) & p. 623-656 (Part II).

Stein, G. (1973). English word formation over two centuries. 7 • E L Q J H Q 1 D U U

Stein G. (1977). English combining-forms. In Linguistica, 9, p. 140-147.

Suffixes and other word-initial elements of English. (1982), Urdang, L. (ed.). Detroit: Gale Research.

Szymanek, B. (2005). The latest trends in English word formation. In Handbook of word formation. Stekauer P., & Lieber, R. (eds). Dordrecht: Springer, p. 429-448.

*The Encyclopædia Britannica: A d*ictionary of arts, sciences, literature and general information. (1859). 8th ed. Edinburgh: Adam & Charles Black. Vol. 17.

Tomaszewicz, E. (2008). Novel words with final combining forms in English. A case for blends in word formation. In P*oznań* studies in contemporary linguistics, 44 (3), p. 363-378.

7 | S H O $ ' L H : R U W E L O G-:|QWVHUQEXFKHQXQG ZLQHO beurteilen ±eine Umfrage zu elexiko. In *Wortbildung im elektronischen Wörterbuch*. Klosa, A. (ed.). 7 • E L Q J H Q p. 197-214.U

Trésor de la langue française informatisé (TLFi) Available at: http://atilf.atilf.fr/

Warner Marien, M. (2006). Photography: A cultural history, 2nd ed. London: Laurence King Publishing.

Warren, B. (1990). The importance of combining forms. In Contemporary morphology. ' U H V V O H U : 8 / X V F K • W] N \ + & 3 I H I I H U Berlin: de Gruyter, p. 111-132.

Wiemeyer, L. (forthcoming). The diachronic productivity of native combining forms in English. In Corpus linguistics, context and culture. Proceedings of the 35th

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

conference of the international computer archive of modern and medieval English (ICAME 35). Nottingham, UK, 30 April–5 May, 2014.

:RUEV ('LH 6SWatergate $IIQUHGHLP :RUWVFKD Intermorphem-gate in den slavischen Sprachen und im Deutschen. In: Sprachtransfer – Kulturtransfer. Salnikow, N. (ed.) Frankfurt/M.: Peter Lang, p. 169–181.

Word parts dictionary. Standard and reverse listings of prefixes, suffixes, roots and combining forms. (2000). Sheehan, M.J. (ed.). 2nd ed. Jefferson, NC & London: MCFarland.

Young, M.E., Norman, G.R. & Humphreys, K.R. (2008). The role of medical language in changing public perceptions of illness. PLoS ONE, 3 (12), e3875. Available at: https://doi.org/10.1371/journal.pone.0003875

Zelle, B.D. (2016). Induction, semantic validation and evaluation of a derivational morphology lexicon for German, PhD dissertation, Heidelberg. Available at: http://www.ub.uni-heidelberg.de/archiv/20539

Zimmerer) $QGUHHYD % 0|EL (2017). Perzeption von DO Sprechgeschwindigkeit und der (nicht nachgewiesene) Einfluss von Surprisal. In Proceedings of the 28th Conference 'Elektronische Sprachsignalverarbeitung 2017' Studientexte zur Sprachkommunikation, 6DDUEU•FNHQ-17 March *HUPDQ\ 2017, p. 174–179.

Contact data		Fields of interest
Dr. Katrin Menzel Postdoctoral researcher Department of Language Science and Technology at Saarland University, Campus A2.2, 66123 6DDUEU•FNHQ e-mail: k.menzel@mx.uni-saarland.de		Corpus linguistics, morphology, discourse analysis, translation studies, historical linguistics

ISSN 2453-8035 DOI: 10.1515/lart-2017-0016

Contact data		Fields of interest
Dr. Stefania Degaetano-Ortlieb Postdoctoral researcher Department of Language Science and Technology Saarland University, Campus A2.2, 66123 6 D D U E U • F N H Q e-mail: s.degaetano@mx.uni-saarland.de		Computational and corpus-based linguistics, data mining, applied linguistics, English scientific language, information density, linguistic encoding, evaluation/appraisal/ stance, translatology

Résumé in English

Our study addresses the diachronic development of combining forms in English scientific texts over approximately 350 years, from the early stages of the first scholarly journals that were published in English to contemporary English scientific publications. Combining forms as bound lexical morphemes (e.g. lith(o)- / -lith, graph(o)- / -graph, bio-, -lysis) share some similarities with affixes and, at the same time, with base lexemes and parts of regular compounds. They seem to have always played a particularly important and productive role among English lexeme-formation elements in languages for special purposes, especially for the creation of new nouns. In this paper, we present a critical discussion of the category of combining forms as well as a case study that examines the role of selected combining forms in English scientific discourse. We use two diachronic corpora that consist of scientific texts from various disciplines from the middle of the 17th century onwards to the beginning of the 21st century – the Royal Society Corpus (RSC) and the Scientific Text Corpus (SciTex). Combining several lexical morphemes within single lexical items is a word formation strategy that is particularly important for informational texts from scientific domains. What we primarily consider is the surprisal value of each unit. Surprisal is an information-theoretic notion related to the predictability and information density of text elements and measures the probability of a unit to occur in a given textual context. We present the insights that can be gained from considering surprisal values of combining

ISSN 2453-8035 DOI:10.1515/lart-2017-0016

forms and the elements, with which they occur in complex lexemes, measuring the probability of morphemes occurring together in specific time periods. The results of our case study have shown that a more predictive and conventionalized use of particular forms allows a more productive use of those forms in closely related or analogous grammatical contexts. Combining forms that are used as components of nouns in a rather predictable way, for instance, become easily productive in other word classes as well.

Key words: combining forms, morphology, history of scientific English, language for specific purposes, information density, corpus linguistics.

Résumé in German

In unserem Beitrag untersuchen wir die Entwicklung von Konfixen (combining forms) in HQJOLVFKHQ :LVVHQVFKDIWVWH[WHQ •EHU HLQHQ seit dem Aufkommen der ersten wissenschaftlichen englischen Zeitschriften bis hin zu]HLWJHQ|VVLVFKHQ $UWLNHOQ DXV HQJOLV Fachzeitschriften. Die von uns betrachteten Konfixe sind gebundene lexikalische Morpheme (z.B. lith(o)- / -lith, graph(o)- / -graph, bio-, -lysis), die einige Gemeinsamkeiten sowohl mit Affixen als auch mit Basislexemen und Bestandteilen YRQ UHJXOlU JHEL OGHWHQ LQ (QJOLVFKHQ EHVRQGHUV EH Neubildung von Nomen in der Wissenschaftssprache schon seit langem eine besondere 5ROOH ,Q GLHVHP %HLWUDJ SUlVHQWLHUHQ ZL Morphemkategorie in Abgrenzung zu anderen Wortbildungselementen. In einer)DOOVWXGLH EHVFKlIWLJHQ ZLU XQV GHWDLOOLHU Wissenschaftsdiskurs. Es werden zwei diachrone Korpora bestehend aus wissenschaftlichen 7H[WHQ YHUVFKLHGHQHU 'LV]LSOLQHQ • -KG EL V $QIDQJ GHV -KG I•U das Royal Society Corpus (RSC) und das Scientific Text Corpus (SciTex). Die Verbindung mehrerer VROFKHU 0RUSKHPH PLW OH[LNDOLVFKHU %HGHXW

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

ist eine wichtige Wortbildungsstrategie in wissenschaftlichen Fachtexten. In der Analyse legen wir einen Schwerpunkt auf den Surprisal-HUW hEHUUDVFKX der jeweiligen Einheiten. Surprisal als Begriff aus der Informationstheorie sagt etwas DXV •EHU GLH 9RUKHUVagbarkeit und den Informationsgehalt von textuellen Bestandteilen und misst die Auftretenswahrscheinlichkeit dieser im jeweiligen textuellen Kontext. Es ZLUG DXIJH]HLJW ZHOFKH (UNHQQWQLVVH DQKDQ Morpheme, mit denen sie in komplexen Lexemen gemeinsam auftreten, gewonnen ZHUGHQ N|QQHQ +LHUEHL ZLUG EHVWLPPW PL untersuchten Morpheme in bestimmten Zeitperioden gemeinsam innerhalb von Lexemen auftreten. Die Ergebnisse zeigen, dass bei einer vorhersagbareren und NRQYHQWLRQDOLVLHUWHUHQ *HEUDXFK EHVWLPPW bzw. analogen grammatischen Kontexten verwendet werden. Konfixe, die als Bestandteile von Nomen vorhersagbarer werden, finden beispielsweise dann auch proGXNWLYHU DOV (OHPHQWH YRQ :|UWHUQ DQGHUH

Stichwörter: Konfixe (combining forms), Morphologie, Geschichte der englischen Wissenschaftssprache, Fachsprache, Informationsdichte, Korpuslinguistik.

Résumé in French (translated by Olivier Landeville)

Notre article a pour objet d'pWXGLHU O'pYROXWLRQ GHV combining forms dans les WH[WHV VFLHQWLILTXHV UpGLJpV HQ DQJODLV GX depuis la publication des premiers journaux scientifiques anglophones jusqu'aux DUWLFOHV FRQWHPSRUDLQV H[WUDLWV GH UHYXHV VRQW GHV PRUSKqPHV (par ex. en anglais lith(o)- / -lith, graph(o)- / -graph, bio-, -lysis) qui ont ceUWDLQHV VLPLODULWpV DYHF OHV DIIL les le[qPHV EDVH HW OHV pOpPHQWV GH FRPSRVp VHPE GHSXLV ORQJWHPSV GDQV OD ODQJXH DQJODLVH formation de nouveaux substantifs dans le langage scientifique. Dans cet article, nous offrons XQ GpEDW FULWLTXH VXU OHV FRQIL[HV HQ W

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

XQH pWXGH GH FDV QRXV H[DPLQRQV OH U{OH scientifique en anglais. Dans le cadre de cette analyse, deux corpus diachroniques FRPSRVpV GH WH[WHV VFLHQWLILTXHV LVVXV GH du 17e VLqFOH DX 19e VLqFOH RQW pWp XWLOLVpV OH 5R\DO Scientific Text Corpus (SciTeX). /D FRPELQDLVRQ GH SOXVLHXU FH W\SH SHUPHWWDQW GH IRUPHU GHV PRWV LQG GH FUpDWLRQ OH[LFDOH LPSRUWDQW a été analysée, l'accent est mis sur les valeurs GH SUpGLFDWLRQ, c'est-à-dire VXU OD SURED pOpPHQW DSSDUDLVVH GDQV XQ FRQWH[WH RX XQH FH FRQFHSW SRUWH VXU GLIIpUHQWV pOpPHQWV 1RXV SUpVHQWRQV OHV FRQFOXVLRQV TXL SHXYH YDOHXUV SRXU OHV FRQIL[HV HW DXWUHV PRU FRQMRLQWHPHQW GDQV GHV OH[qPHV FRPSOH[HV OD TXHOOH PHWV DX[TXHVWLRQV DSSDUDLVVHQW GDQV G VSpFLILTXHV /HV UpVXOWDWV GH QRWUH pWXGH IRUPHV HVW SOXV SUpYLVLEOH FH TXL HQW plus productive dans des contextes grammaticalement similaires et analogues. Les FRQIL[HV XWLOLVpV SDU H[HPSOH HQ WDQW TXH SUpYLVLEOH GHYLHQQHQW IDFLOHPHQW SURGXFWL

Mots-**clés:** confixes (combining forms), morphologie, histoire de l'anglais scientifique, OD QJXH GHV VFLHQFHV LQIRUPDWLRQ, linguistique de corpus

Résumé in Russian (translated by Ekaterina Lapshinova-Koltunski)

< ^Zgghc klZlv_ jZkkfZljb\Z_lky bklhjbq_kdkehj`Zau\blb_ keh\ \ Zg]ebckdbo gZmqguo l_dklZo gZ ijhly`_gbb ijl gZqbgZy hl jZggbo gZmqguo im[ebdZpbc gZ Zg]ebc kh\j_f_gguf Zg]ebckdbf Dhfihg_glu keh`guo keh\ \k_]^ b ijh^mdlb\gufb ijhp_kkZfbgbyhjfebjdlZq_kdbo _^bg

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016

ki_pbZevguo yaudZo hkh[_ggh \ nhjfbjh\Zgbb gh\uo
^Zgghc klZlv_ ij_^klZ\e_g djblbq_kdbc ZgZeba dZl_]hj
keh\ gZ ijbf_j_ ba[jZgguo dhfihg_glh\ \ Zg]ebckdhf gZ
bkihevam_fbZ\oZjhgguo dhjimkZ gZmqguo l_dklh\ ba jZ
kha^Zgguo gZ ijhly`_gbb fgh]bo-]e_b aZdZggZy\ZkdhfZ
± Royal Society Corpus (RSC Dhjimk Dhjhe_\kdh]h GZmqgh]h
Scientific Text Corpus (SciTex dhjimk gZmqguo l_dklh\ gZ Zg]eb
kha^Zgguo \ >ZjfrlZ^l_ b KZZj[jxd_g_ >ey bgnhjfZlb\g
^bkdmjkZ oZjZdl_jguf kihkh[hf keh\hh[jZah\Zgby y\ey
e_dkbq_kdbo fhjn_f \ h^gm e_dkbq_kdmx _^bggZbkm
bgl_j_ksurprisal(g_ ij_^kdZam_dfZ\lwc _^bgbpu lh _klv _j
ihy\e_gby ^Zgghc _^bgbpu \ ^Zgghf dhgl_dkl_ Fu ij_
bkke_^h\Zgby _ephsabdhufihg_glh\ keh`guo keh\ b we_
dhlhjufb hgb khq_lZxlkgyuo\ ek_edhk`_fZo baf_jyy _jh
khq_lZ_fhklb fhjn_f \ hij_^_ezggu_ i_jbh^u \j_f_gb J
bkke_^h\Zgby ihdZau\Zxl qlh [he__ ij_^kdZam_fh
bkihevah\Zgb_ hij_^_ezgguo nhjf iha\hey_l [he_
bkihevah\Zgb_ wllb_kgmhjf\yaZgguo dhgl_dklZo

Ключевые слова: dhfihg_glu keh`guo keh\ fhjnheh]by
Zg]ebckdh]h yaudZ Yaud ^ey klSPbZeBvgmhjfZpp_beh_cgZ
iehlghklv dhjimkgZy ebg]\bklbdZ

Article was received by the editorial board 27.06.17.
Reviewed 23.07.17. and 21.09.17.
Similarity Index 3%

 ISSN 2453-8035 DOI:10.1515/lart-2017-0016