Astronaut, *astrology*, *astrophysics*: About Combining Forms, Classical Compounds and Affixoids

Dieter Kastovsky
Vienna University

1. Introduction[1]

About seventy percent of the English vocabulary is loans, the majority coming from French, Latin, Greek and what has been called Neo-Latin, the *lingua franca* of the Middle Ages and the Renaissance, which was a mixture of Latin and Greek vocabulary. This massive borrowing in Middle and Early Modern English resulted in the adoption of numerous foreign word-formation patterns. In the classical handbook on English word-formation, Marchand (1969), 56 of the 66 prefixes listed are foreign, and 48 of the 81 suffixes treated are foreign as well. Beside prefixations and suffixations, which have been investigated fairly well and whose morphological status is relatively clear, there are also numerous formations such as

(1) a. *astronaut, cosmonaut, Marsnaut, astrophysics, astro-bio (expert), biosphere, noosphere, agro-chemical, agriculture, patricide, infanticide, galvanoscope, Ameriphile, cryosleep ethnohistory, television, telepresence, holomovie;*
    b. *hyperactive, hyperaemia, omnipresent, omnisphere, omniscient, supernatural, extra-ordinary, exocentric, exo-biology, multiparous;*
    c. *diamondiferous, hepatitis, wake-up-itis, Winston-Churchill-itis, astrology, phonology, biology, monologue, polylogue, polygeny, megalomania, star-mania, geologist, aerologist, geological, selenological,* etc.

The morphological status of these formations and their constituents is less obvious. They contain so-called "combining forms" and are not dealt with in Marchand (1969), save for some exceptional cases such as *hyper-* or *extra-* and a few others. The examples in (1) are partly established lexical items listed in dictionaries and partly neologisms collected from science fiction novels, the latter demonstrating the productivity of the formative processes involved, not simply in technical and scientific terminology, where they are abundant. In view of the increased frequency of such formations, it is surprising that so far they have never been investigated systematically. This had already been pointed out by Stein (1977: 140), who noted that "none of the more than 1800 publications on English word-formation which are listed in my bibliography [Stein 1973] deals with this problem in detail". And even though a number of publications have appeared in the meantime, e.g., Bauer (1998), Lehrer (1998), Warren (1990), and Prćić (2005, 2007, 2008), the situation has not really changed, as Prćić states in the latest instalment of his contributions to this area, which deals with the demarcation between final combining forms and suffixes:

> [M]odern morphological theory has still not worked out a principled and consistent way of distinguishing between affixes and combining forms in general, and between suffixes and final combining forms in particular. […] This unsettled state of affairs has had adverse implications not only for the overall theory of word formation in English, but also for lexicographic methodology and practice as well as for language teaching (Prćić 2008: 2).

This is certainly true, and the following contribution will try to make up a little for this neglect. But not being a lexicographer nor a language didactician, I will not deal with these aspects in what follows, but will restrict myself to some remarks concerning the status of these formations from the point of view of word-formation theory, hoping that this will contribute in the long run to the practical application in dictionary-making and language teaching as well. I am, on the other hand, aware of the fact that these practical applications may sometimes have to gloss over certain distinctions that the theory-minded lexicologist would want to introduce, or, as Bernard Quemada, one of the fathers of the *Trésor de la language française*, once said to me, when, as a young student and Coseriu-trained semanticist at the University of Besançon in 1966, I was arguing with him about a lexical definition: "Monsieur, vous êtes lexicologue, mais moi, je suis lexicographe, ce qui est une chose tout-à-fait différente".

The issue addressed in the following is the morphological status of these "combining forms", which are central for the type of word-formation illustrated in (1). I will therefore first look at some definitions of these "combining forms" and the criteria by means of which a distinction might be established between them and stems/roots on the one hand, and affixes on the other. I will then try to show that these attempts are rather problematic in view of the shaky theoretical basis underlying the notion of "combining form", and I will finally propose to abandon the notion altogether, since modern word-formation theory can well do without it.

2. Combining forms: definitions

The term "combining form" seems to go back to the *Oxford English Dictionary*, or, more precisely, the *New English Dictionary* (1884–1928), as it was then called. There the term was used to label parts of loans from Greek or Latin, or English formations using constituents which were neither words proper nor easily identifiable as affixes, cf.

(2) *Astro-*[…] repr. Gr. *άστρο-* **stem and comb. form** of *άστρον* 'star'…
Bio- […] repr. Gr. *βίο-* **stem and comb. form** of *βίος* 'life, course or way of living'…
Neo- **combining form** of Gr. *νίος*, new […] common in recent use as a **prefix** to ads. and sbs.
Micro- […], before a vowel *micr-,* repr. Gr. *μίκρο-,* **combining form** of *μίκρος* small, used chiefly in scientific terms.[2]

Even these random examples of some entries in the *OED* show that there was no systematic principle behind the classification of such elements as either stems, combining forms or affixes; cf., the dual affiliation assignments with *astro-, bio-* and *neo-*. This might be compared with Marchand's 1969 treatment, to which I will return below: there, neither *astro-* nor *bio-* show up in any form, whereas *micro-* and *neo-* are listed as prefixes. Now all of them are regarded as "combining forms" by the *OED*, which would therefore accord them equal status, but Marchand is much more selective and includes only *micro-* and *neo-* under the heading of prefixes, rejecting the notion of "combining form" altogether, a problem to which I will return below. At the inception of the *NED*, however, morphological theory was in its infancy and, moreover, the original *OED* (*NED*) edition did not even contain any definition of "combining form" and consequently had no criteria by means of which it could be distinguished from other lexical elements such as words or affixes.

It was only in the later editions of the *OED* that some definition was supplied under the entry *combining form*:

In Latin and other languages, many words have a special combining form which appears only in compounds (or only in compounds and derivatives). […] The foreign-learned part of the English vocabulary also shows a number of special combining forms; cf. *electro-,* combining form of *electric*, in such compounds as *electromagnet.* (*OED*).

---

[2] Emphasis added by the author.

This quotation is from Bloch and Trager (1942), a widely-used introduction to linguistics in the U.S. at the time, which apparently describes the practical *OED* use of the term, but does not really provide specific criteria for its delimitation. It is noteworthy that the *OED* use itself is older, but that the *OED* does not refer to it, although it would normally provide the earliest quotation of the entry in question.

In the *Shorter Oxford English Dictionary* we find the following entry, also without any more specific criteria for delimitation and identification: "a form of a word used (only) in compounds, as *Indo-* repr. *Indian* in *Indo-European*" (*Shorter Oxford English Dictionary*, s.v. *combining,* Comb. *combining form*).

The definition in Webster (1994) is somewhat more restrictive than the *NED/OED* one, but still lacks adequate criteria for the delimitation between affixes and combining forms:

> A linguistic form that occurs only in compounds or derivates and can be distinguished descriptively from an affix by its ability to occur as one immediate constituent of a form whose only other immediate constituent is an affix (as *cephal, cephalic*) or by its being an allomorph of a morpheme that may occur alone as *electro* representing *electric* in *electromagnet* […] or can historically be distinguished by the fact that it is borrowed from another language in which it is descriptively a word or a combining form (*Webster* 1994: s.v. *combining form*).

Thus, it is not clear whether *neo- micro-* would qualify as combining forms or affixes, and the same applies to *-(o-)logy*, which is also ambivalent between the categories of combining form and suffix.

3. Problems with combining forms

3.1. The basic problem with these definitions is that the linguistic status of these elements was never really made clear in the *NED*, nor were there any criteria by means of which they could be distinguished from other types of lexical element such as words, roots, stems or affixes. But then neither were these elements really well-defined within a morphological theory at the time, which was only developed much later than the origin of the *OED*, and even today terminology still varies considerably in this respect.

The doyen of English word-formation, Hans Marchand, had of course noticed this problem in his handbook (Marchand 1969) and criticised the unsystematic *OED* practice, but he was, unfortunately, far from consistent either. Thus, with regard to morphemes such as *ante-, extra, intra-, hyper-, para-, multi-,* etc. he says:

> The patterns of coining are Neo-Latin (though some of the types are already Old Greek or Ancient Latin). With these **particles** [my emphasis] there is a practical difficulty. They may represent 1) such elements as are prefixes […] in Latin or Greek, as **a-** (*acaudal*, etc.), **semi-** (*semi-annual*), 2) such elements as exist as prepositions or particles with an independent word existence, as **intra, circum / hyper, para**, 3) such as are stems of full words in Latin or Greek, as **multi-, omni- / astro-, hydro-.** This last group is usually termed 'combining forms' (OED, Webster). In principle the three groups are on the same footing from the point of view of English word-formation, as they represent loan elements in English with no independent existence of words. That **macro- micro-,** and others should be termed combining forms while **hyper-, hypo-, intro-, intra-** and others are called prefixes by the OED, is by no means justified (Marchand 1969: 131–132).

This criticism is of course appropriate, but I think that in his conclusion Marchand is too restrictive and also inconsistent. Thus he argues:

> Only such particles as are prefixed to full English words of general, learned, scientific or technical character can be termed prefixes. […] We cannot, however, undertake to deal with all the prepositive elements occurring in English. Such elements as **astro-, electro-, galato-hepato-, oscheo-**, and countless others which are used in scientific or technical terminology

have not been treated in this book. They are of a purely dictionary interest in any case. In the main, only those particles have been considered that fall under the above groups 1) and 2). But we have also included a few prefixes which lie outside this scope, as prefixes denoting number (**poly-, multi-**), the pronominal stem **auto-** […] and particles which are type-forming with English words of wider currency (as **crypto-, neo-, pseudo-**).

Thus he criticises the *OED* for treating elements such as *a-, semi-* (= original prefixes), *hyper-, intra-, para-* (= original particles) and *astro-, electro-, hepato-, multi-, omni-* (= original stems) as different categories, as affixes and combining forms, although from a Modern English point of view they should have the same status, foreign bound forms without independent existence in English. But then his practical solution is just as inconsistent. Thus he regards as prefixes only such particles as are prefixed to "full English words". And he excludes elements such as *astro-, electro-, galato-, hepato-, oscheo-,* etc. from treatment in his book because they are of a "purely dictionary interest" — a very questionable argument in view of their important role in scientific terminology, which could not do without them. But then he includes among his prefixes elements such as *poly-, multi-, auto-, crypto-, neo-,* originally a mixed bag of prefixes, stems and particles subsumed under the label of "combining forms" in the *OED*, which in view of his own statements is far from consistent.

A similar problematic decision is made by Marchand concerning the delimitation of suffixation and combinations containing *OED's* "terminal elements", the word-final mirror image of the "combining forms". Thus, in connection with the word *galvanoscope* he says:

> A word such as *galvanoscope* […] is either not analysed at all or said to be *galvano-* [listed as a combining form of *galvanic* in *ShOED*, DK] plus *-scope*. But what is *-scope*? The OED terms elements like *galvano-* 'Combining forms' and elements such as *-scope* 'terminal elements'. This terminology only begs the question as to what these elements really are in word-formation. […] Neither *scientist* nor *galvanoscope* are analysable as 'English word plus affix'. Yet there is a great structural difference between the two words. The radical of *scientist* is immediately connected with the word *science* of which it is merely an allomorph, so to speak. The case is different with *galvanoscope*, and, generally speaking, with combinations with 'terminal elements'. The first-word cannot be connected with any independent English word as its allomorph. I have therefore treated words of the type *scientist* while I have left out combinations with 'terminal elements' (Marchand 1969: 218).

What Marchand has overlooked is that *scient-, galvan-* (derived from the name *Galvani*) and *-scope* have the same status: neither are suffixes but stems, i.e., bound lexical bases, and suffixation does not play any role in this analysis. In other words, Marchand's main criterion is whether these elements — prefixes or suffixes — can "be prefixed [or suffixed, D.K.] to **full words** without, however being **independent words** in English" (Marchand 1969: 132 [my emphasis]). It is this exclusively word-based analysis which creates the classification problem, to which I will return below, proposing a solution that allows for stem-based processes in English. In this connection, another major problem will be addressed, the demarcation between words, stems, and affixes on the one hand and elements which are neither on the other, viz. curtailed words and stems (clippings) and parts of blends, so-called splinters (cf. Lehrer 1998), which also play a role in the demarcation of combining forms.

Similar problems also arise in Stein (1978), Bauer (1983), and Plag (2000), who recognise the problems involved in Marchand's treatment of these formations, but do not question the legitimacy of the categories "combining form" and "terminal element". It should be added that this distinction between combining forms and terminal elements has now been replaced by the distinction between "initial combining forms" (ICF) and "final combining forms" (FCF), introduced by Laurie Bauer (1983) in view of the similarity between the two. Bauer himself in a later paper (Bauer 1998) voices some doubts as to the legitimacy of these elements, however. Although basically accepting these categories as such, he views them as having something in common with compounding on the one hand, and blends on the other, a position that I will extend below to a scale of free and dependent elements in word-formation, which will make the category of "combining form" superfluous.

3.2. Before I turn to my proposal to abandon this category altogether, however, the most recent serious attempt to come to terms with it deserves a more detailed comment because, despite its merits, it underlines the inherent flaws of the category as such. In a number of papers, Prćić (2005, 2007, 2008) has tried to

(1) offer one method of drawing a systematic dividing line between prefixes and ICFs, by putting forward an ordered set of shared and distinguishing criteria, based on formal, functional, semantic and pragmatic properties of prefixes and ICFs; and (2) to define the categories of prototypical prefix and prototypical ICF, which would, consequently, help to assign — and re-assign — each bound initial lexical element to one of the two categories in a synchronically more justified and appropriate way (Prćić 2005: 314).

A similar attempt is outlined in Prćić (2008) with regard to the delimitation of final combining forms (FCF) and suffixation. In the following, I will concentrate on the attempted demarcation of prefixes and ICFs; since the same kind of analysis is also applied to the demarcation of suffixes and FCFs, the arguments presented in this section with regard to ICFs also hold for FCFs and will therefore not be repeated.

Prćić regards both prefixes and combining forms as signs with an identifiable form/meaning relation in the Saussurean tradition, which is certainly appropriate. They are, moreover, bound elements and require companions (right-hand ones in the case of prefixes/ICFs, or left-hand ones in the case of suffixes/FCFs). These companion elements can be either free forms (words), e.g., *geo-chemistry* or final combining forms, e.g., *geo-logy*, but not suffixes, e.g., **geo-ist*. Note that Prćić does not treat *-(o)logy* as a suffix, but as a combining form, which already demonstrates the fluidity of the concept, since *-logy* would strike one much more as suffix-like than as resembling a lexeme, which would be the next-of-kin category of combining forms according to his demarcation. Prototypically, the combination of right-hand input elements (= determinata in Marchand's terminology) and prefixes or ICFs produces binary endocentric structures, as word-formations usually do. Besides these shared properties, Prćić lists eight features which distinguish prefixes and ICFs from each other:

(3) 1. Category membership, 2. Distinctive form, 3. Co-occurrence restrictions, 4. Syntactic function, 5. Head-modifier relation, 6. Semantic meaning, 7. Morphosemantic patterning, 8. Productivity.

It should be pointed out, however, that Prćić admits at the outset that there is no hard and fast boundary between prefixes and ICFs, so that these criteria always apply to a greater or lesser degree. In the following I will try to show that the criteria are indeed not really helpful to establish combining forms as a category of their own, although they do highlight certain property differences characterising the constituents in question.

Criterion (1), category membership of the initial elements, postulates that prefixes belong to a (relatively) closed set of lexico-grammatical units, into which new members are rarely admitted, but that ICFs belong to a (relatively) open set of lexico-grammatical units, into which new members are fairly readily admitted. This criterion is dubious on two counts: it cannot be quantified — note the qualification "relatively" — and it is disproved by historical evidence, since the prefix category allows new members by borrowing and by the shift from independent lexical items to prefix-status, e.g., borrowed *a-, de-, dis-, en-, in-* and many others, and shifted native items like *fore-, mid-, out-, under-,* or *-monger, -wards, -wise*, etc.

Criterion (2), distinctive form, claims that prefixes have no distinctive form, whereas ICFs end in a vowel (*-o-* as in *anthropo-logy, -i-* as in *agri-culture*, etc.). This criterion is also problematic, since many prefixes, such as *a-, ante-, anti-, auto-, be-, bi-, co-, de-, demi-, di-, epi-, fore-, hypo-, mono-, multi-, pre-, pro-* and others also end in a vowel, and the status of the final vowel in the ICFs is questionable. Thus it is by no means clear whether the vowel in *concav-o-convex, insect-i-cide, phot-o-graphy,* let alone *speed-o-meter, hel-i-port* really belongs to one of the constituent members of the formation, as is sometimes claimed, or whether it is merely a "linking vowel" or *Fugenelement*, triggered by one of the constituents, and therefore belongs to neither constituent but is a morphological

element of its own. Historically, the vowel in these combining form structures almost always represents a stem-formative of the first member, but this had lost its function in the same way as the original case endings in German compounds like *Frau-en-kirche, Grab-es-stille,* where we now also find linking elements that would not be grammatically justified, such as *-s-* with feminines as in *Liebe-s-trank, Leistung-s-druck, Pension-s-kürzung,* etc. Thus, the presence of a linking element might just as well be triggered by the corresponding first or second element on the basis of some historical analogy, with the linking element not being part of either constituent. A very good — but controversial — example is *-(o)logy*. In *anthrop-o-logy*, *cosm-o-logy*, the *-o-* historically represents an original stem-formative. This is also true of *bio-logy, geo-logy;* the combining form or suffix (whichever it is) might therefore be analysed as *-logy*, since here *-o-* is clearly (and historically) part of the first element. But in *Kremlin-o-logy, hamburger-o-logy, life-o-logy*, quoted in Prćić (2005), the suffix either seems to have adopted the form *-ology*, which could then be treated as a morphologically conditioned allomorph of *-logy*, or the suffix triggers the insertion of a linking element, which does not belong to either constituent. To me, the latter solution seems to be more plausible. By the way, since linking elements also occur in regular English compounds, viz. *-s-* as in *officer's mess, driver's license,* etc., we cannot claim that combining forms end in a particular vowel or set of vowels.

Criterion (3), co-occurrence restrictions, deals with the type of element which can be combined with what other type of element. According to Prćić, prefixes only occur with free forms, but not with suffixes or FCFs, e.g., **hyper-ize, *co-phobia, *mis-phagous*. ICFs, on the other hand, can occur with free forms, e.g., *agri-culture*, or FCFs, e.g., *morph-o-logy*, but not with suffixes, e.g., **lexic-o-hood*. This, however, means that *-o-logy* cannot be treated as a suffix but must be regarded as a FCF, which is somewhat problematic, as we have seen in connection with criterion (2).

Criteria (4), syntactic function and (5) head-modifier relation, are not really different and refer to the type of endocentric relation between the right-hand (= head/determinatum) and the left-hand (= modifier/determinant) constituents. Prefixes are supposed to enter only "into subordinative endocentric relations with their heads, because in the output prefixation the modifier (prefix) carries less structural and semantic weight than the head (base), as in *re-write*" (Prćić 2005: 322). Prćić mentions *en-* (e.g., *encage*), *be-* (e.g., *becalm),* *a-* (e.g., ablaze) as exceptions because these are allegedly class-changing prefixes, but this is questionable; cf. Kastovsky (1986), where it is shown that they fit into the normal prefixal role as modifiers if we analyse the second part as zero-derivatives. ICFs, on the other hand

> enter into coordinative endocentric relations with their heads, because in the output composition the head (FCF or base) and its modifier (ICF) carry roughly equal structural and semantic weight, as in *morpho-logy*. This property of ICFs is comparable to that of free-standing bases in composition (Prćić 2005: 322).

The nature of this criterion is far from clear. First of all, any word-formation follows the modifier/head schema (at least in this order in English), which makes the head the dominant element, no matter whether it is a lexeme (word) or a suffix. Thus all genuine compounds, prefixations and suffixations are endocentric from this point of view. Secondly, the notion of "coordinative endocentric relation" is a contradiction in terms: endocentric constructions necessarily have a head, coordinative ones do not and therefore cannot be endocentric. In fact, what Prćić seems to have in mind here is actually something like "degree of semantic density", where the prefix is more general, semantically "less dense", and the lexical head (and FCF) as well as an ICF are semantically "richer", whatever this means. This question, however, is actually dealt with in connection with his next criterion.

Criterion (6), semantic meaning, involves a distinction between prefixes, which are said to be characterised by functional/lexical meaning of varying semantic density equivalent to that of prepositions, adverbs, numerals, adjectives and verbs, and ICFs, which typically have lexical meaning of a high semantic density, equivalent to that of nouns or adjectives. Apart from the fact that both overlap with regard to adjectival meaning, the notion of "semantic density" is again relative and not really quantifiable and Prćić provides no definition or criteria to measure "semantic density". The same criterion is also suggested for the distinction between suffixes and FCFs, but again suffixes may be very specific, cf. *-age* in *vicar-age* 'abode of a vicar', *anchor-age* 'fee for anchoring' etc. Thus, at best the criterion again points towards a scale rather than an all-or-none distinction between affixes and CFs,

and, moreover, would presuppose a much more elaborate word-formation semantics than we have at the moment in order to specify what is meant by "semantic density".

Criterion (7), morphosemantic patterning, involves recurrent, ready-made, automatic morphosemantic patterning in the case of affixes, whereas with CF there may be variation, as with compounds. But again, this is a cline, as is also admitted by Prćić, since first or second members of compounds, such as *bio-* or *man*, can form series, and, as is proved by suffixes such as *-hood* or *-dom* or prefixes such as *fore-, out-*, lexemes can develop into affixes via such series.

Criterion (8), productivity, refers to the systematic and pattern-conforming readiness to combine with bases, which is allegedly high with affixes, whereas combinations involving CFs are far less productive. But again, this is a matter of degree, since some CFs, e.g., *bio-* or *-logy* (if this is treated as a FCF), are just as pattern-forming as affixes.

On the basis of the interaction between these 8 criteria, Prćić postulates prototypical synchronic prefixes and suffixes as well as prototypical synchronic ICFs and FCFs, cf.

(4) a. Prototypical synchronic prefixes: *anti-(aging), co-(produce), de-(stabilize), dis-(connect), e-(cash), en-(large), ex-(president), hyper-(active), in-(audible), inter-(national), mis-(spell), multi-(national), non-(academic), out-(play), over-(cook), post-(Elizabethan), pro-(British), re-(write), self-(esteem), sub-(conscious), super-(fine), trans-(atlantic), ultra-(modern), un-(just), under-(estimate)*

b. Prototypical synchronic ICFs: *aero-(dynamics), Anglo-(-phile), audio-(-metry), biblio-(-graphy), chrono-(meter), dactylo-(-scopy), geo-(chemistry), hydro-(-pathy), morpho-(-logy), neuro-(anatomy), ophthalmo-(-logy), phono-(-tactic), physio-(therapy), tacho-(-gram), xeno-(phobia), agri-(culture), alti-(meter), calli-(-graphy), denti-(-form), fungi-(-vorous)* (= classical ICFs); *filmo-(-graphy), kisso-(-gram), heli-(port), insecti-(-cide)* (= modern ICFs)

c. Prototypical synchronic suffixes: *(wash)-able, (magic)-al, (contest)-ant, (bankrupt)-cy, (star)-dom, (employ)-ee, (mountain)-eer, (wide)-en, (teach)-er, (care)-ful, (child)-hood, (atom)-ic, (solid)-ify, (connect)-ion, (yellow(-ish), (violion)-ist, (attract)-ive, (slow)-ly, (enjoy)-ment, (kind)-ness, (invent)-or, (poison)-ous, (friend)-ship, (east)-ward, (milk)-y*

d. Prototypical synchronic FCFs: *(tachy-)-cardia, (mono-)-chromatic, (fungi-)-cide, (demo-)-cracy, (poly-)-gamy, (photo-)-graphy, (anglo-)-mania, (baro-)-metric, (gastro-)-nomy, (taut[o])-onym, (osteo-)-pathy, (ichtyo-)-phagy, (xeno-)-phobia, (logo-)-rhea, (stereo-)-scopy* (= classical FCFs); *(Monica)-gate, (kisso-)-gram, (milka-)-holic, (webo-)-pedia, (city)-scape, (swima-)-thon, (six)-pounder, (dark)-haired, (horror)-meister, (dulls-)-ville* (= modern FCFs).

Prćić's analysis is certainly an important contribution to the description of this rather neglected area, but it also raises a number of problems in connection with the application of the criteria used and the results arrived at, which confirms the suspicion that the category "combining form" itself is of doubtful validity.

One problematic aspect is the introduction of modern, i.e., English-based CFs such as *filmo-(-graphy), kisso-(-gram), (Monica)-gate*. The status of *filmo-, kisso-* is extremely dubious. Prćić treats *-o-* as part of the first constituent, but this is questionable. His analysis forces him to assume that the ICFs *filmo-* or *kisso* have to be regarded as allomorphs of *film, kiss*, respectively, which strikes me as rather counterintuitive. As I have already pointed out, in such instances, and with the traditional-latinate CFs, it would be preferable to treat *-o-* here as a real linking element, not belonging to any of the constituents, even though historically this started out as stem-formatives of the first constituent with the classical ICFs, from where it was analogically extended to the modern forms. I would therefore suggest that the morpheme *-gram* simply triggers the insertion of a linking vowel, unless the first element already ends in one, just as *-logy* does, i.e., there are no ICFs *filmo-, kisso-*. As to *-gate,* its status is similar to *-burger*, i.e., it is a reinterpretation of the second constituent of a compound (or apparent compound), by which it was converted into a suffix or semi-suffix, but certainly not into a CF. And just as *burger* has become a separate lexical item, the same might happen to *gate* in the sense of 'political scandal'. It just so happens that compound members become more and more suffix-like, cf. Marchand's

term "semi-suffix" or the alternative term "affixoid", which can be illustrated by the examples *fish-monger, back-wards, clock-wise*.

Another problematic case is the treatment of *self* as a prefix, which in Marchand (1969: 78 f., 87 f.) is dealt with as part of compounding, and rightly so, because *self* is an independent lexical item, and not a combining form. The treatment of *e-* in *e-cash* 'electronic cash' as a prefix also definitely goes against Prćić's semantic criterion for prefixes, which are supposed to be semantically general, whereas *e-* is semantically very specific, since it is a clipping of *electronic*.

One important aspect that emerges from Prćić's analysis, however, is that there is a gradience between combining forms and affixes, cf. the classification problems just mentioned. On the other hand, it is exactly this aspect which makes Prćić's analysis less attractive than it might be otherwise, because he restricts this cline to the gradation between affixes and CFs, but does not include the more general cline that exists between words and affixes in general because of the historical shift of first or second compound members to affixhood, cf. the history of *-dom* or *-hood*, or the semantic development of the locative particles *out, over, under* in *out-bid, overstate, underestimate* to elements that express degree. It would seem that the cline between CF and affix is only a part of this more general scale, which, as we will see below, extends to other domains such as clipping and blending, which have to be included as well in this scale of decreasing independence of constituents.

4. A solution: morphological typology and the scale of boundness

4.1. Despite a number of positive aspects, Prćić's approach suffers from the same deficiency that characterizes the other approaches using the notion of "combining form". Despite the attempt to make it more precise by postulating a number of intersecting criteria, the category still remains vague and covers a range of different though overlapping phenomena that ultimately also include clippings, clipping compounds, emerging affixes, and might even be extended to blending. One reason is that the CFs go back to a number of different sources with distinct properties, so that it is not surprising that synchronically this category would be extremely heterogeneous and is therefore highly questionable as an analytic device. The other reason is that all these approaches, although aware of the heterogeneity of the English vocabulary because of its massive borrowing, do not really accept as a fact that the English morphological system itself is heterogeneous. In other words, the input to morphological processes is not homogeneous (cf., e.g., Kastovsky 1992, 1994, 2006). This conclusion is based on the assumption that the morphology of a language, among other things, is characterised by its input, which can differ from language to language and need not be homogeneous within one and the same language. In order to deal with this problem, it is necessary to distinguish three types of input to morphological processes, which characterise three types of morphology, neither of which need necessarily occur in a pure manifestation. Rather, more often than not, we find a mixed type. If one accepts this premise, the category of CFs is superfluous and can be discarded, because the formations concerned can be taken care of by other morphological categories which will be needed anyway.

4.2. In this framework, the central notion is the "lexeme", which can have different types of representation. The term "lexeme" has now been adopted by many European morphologists instead of "word" as a cover term for more specific lexical representations because of the ambivalence of the latter term. Following Matthews (1974) and Lyons (1977), it can be defined as follows:

(5) **Lexeme**: a simple or complex dictionary entry, i.e. an abstract lexico-semantic entity as part of the lexicon of a language, e.g. *WRITE, WRITER, REWRITE, BEARD, BEARDLESS, UNBEARDED, HELP, HELPER, HELPLESS, BIRD, BIRDLIKE, BIRDCAGE*, etc.

Therefore, what is usually called word-formation in actual fact deals with lexeme-formation. Lexemes allow three different prototypical representations that enter into the various morphological processes that operate within inflectional morphology and word-formation (= lexeme-formation):

(6) **Word**: an independent, meaningful syntactic element, susceptible of transposition in sentences; it may be simple or complex, and is thus the concrete realisation of a word-form (inflectional form) in an utterance, e.g. *man, men, go, go-es, mad, madman, writ-er, writer-s, or, the*, etc., i.e., it is a free form.

(7) **Stem**: a word-class specific lexeme representation which cannot occur on its own as a word but has to combine with additional derivational and/or inflectional morphemes to function as a word, i.e., it is a bound form. It may itself contain derivational affixes or so-called stem-formatives, which determine the inflectional category, e.g., G $bind_V$-(*-en, -e, -est,* etc.), OE luf_V-(*-ian, -ie, -ast, -od-e,* etc.), luf_V-$estr_N$-(e) 'female lover', Mod. E. $scient_N$-(*-ist*) vs. *science,* $dramat_N$-(*-ic*) vs. *drama, bio-, astro-, -naut,* Lat. am-$ā_V$-(re), hab-$ē_V$-(re), where $bind_V$-, luf_V-, luf_V-$estr_N$, $scient_N$-, $dramat_N$-, am-$ā_V$-, hab-$ē_V$-, $astr_N$-, $naut_N$- are stems.

(8) **Root**: an element that is left over when all derivational, stem-forming and inflectional elements are stripped away, and which needs morphological material to become a stem, to which further morphological material has to be added for the stem to become a word. Such roots can either be affiliated to a particular word-class, or can be word-class neutral. In this case, the word-class affiliation is added by a word-formative process, cf. IE roots like *gVn- (the source of L *genus, gignere, cognatus,* E *kind, cognate*)*, *mVd- 'measure' (cf. OE *metan*)*, *CVd- 'eat' (cf. OE *etan*)*, *wVr- 'turn' (cf. Lat *uer-t-ere, uer-m,* OE *weor-þ-an, wyr-m*, etc.)*, or Semitic *k.t.b.* 'write*'*.

This results in a distinction between word-based, stem-based and root-based morphology. The terms "stem" and "root" are used here in a very specific way, which is geared to the history of the IE languages. Indo-European was root-based at a very early stage and needed morphological material to build stems — the so-called stem-formatives — to which then the actual inflectional nominal or verbal endings were added to form words. In the later stages of the IE languages, root-based morphology was replaced by stem-based morphology, since the stem-formatives were gradually lost as distinctive morphological entities, merging with the roots or the inflectional endings. This resulted in a system that contained bound stems followed by inflectional endings, a situation that characterised the older Germanic languages (cf. Kastovsky 1992, 1996, 2006), but also other IE languages such as Latin. At this stage, the original distinction between stems and roots had lost its relevance, and inflectional and derivational morphology became stem-based. But the development did not stop there. Owing to the progressive loss of inflectional endings, the Germanic languages, and English in particular, gradually developed a word-based morphology as well. This dominates in English inflection, whereas German has preserved a mixture of stem-based verbal and word-based nominal and adjectival inflection. But, as will be demonstrated below, besides the native word-based type of word-formation, English has again adopted stem-based word-formation at the level of word-formation on a non-native level, type *scient-ist, dramat-ic* etc. as a result of borrowing.

Unfortunately, there is quite a bit of terminological confusion in this respect, since the terms "stem" and "root" are often used interchangeably or in a different sense from that used here. Thus Bauer (1983: 20–21; 1992: 252–253) distinguishes between roots as forms which are not further analysable and which remain when all derivational and inflectional elements have been removed, and stems (only relevant in inflection), which remain when the inflectional morphemes have been removed. In his analysis, both roots and stems can be both free or bound. Thus, in *un-touch-able-s, touch* is the root, and *untouchable* the stem. Giegerich (1999) distinguishes between words and roots, with roots being basically equivalent to what I call stems, and Huddleston and Pullum (2002: 1624–1625) distinguish between free and bound bases. On the other hand, many linguists use the two terms "root" and "stem" non-distinctively, a practice which should, however, be avoided in view of the necessary distinction between these two categories for early Indo-European. In the following, I will use the term "stem" for what some linguists might prefer to call "root" also with reference to English or German, i.e., for a lexeme representation that is bound and can only occur as a word with additional morphological material, which can be either inflectional or derivational.

4.3. Old English inflection was both stem- and word-based, because of its still fairly rich inflectional system, but in Modern English, inflectional morphology is purely word-based (save for some foreign plurals such as *alumn-i, fung-i, phenomen-a, appendic-es,* etc.). This is due to the drastic loss of inflection since OE, which led to the emergence of an inflectionally unmarked base form. This form can be used as a word in the above definition (6) without any morphological material having to be added. This also holds for word-formation on a native basis of coining, as Marchand (1969: 5ff.) calls it, both as regards compounding and derivation, cf. *book, bookish, booking, bookstore.* This means that this domain of word-formation is word-based, and it is this situation that Marchand has somewhat over-generalised.

The influx of French and in particular Latin and Greek loans in late Middle English and Early Modern English led to the emergence of new word-formation patterns based on loans which had been derivationally connected in the source languages. But both inflectional and derivational morphology in the source languages had been stem-based. Therefore the adoption of these non-native word-formation patterns led to the re-introduction of stem-based derivational morphology except for those instances where the new patterns had been fully nativised. In some cases, the patterns operate at both levels at the same time. Therefore, no clear-cut boundary exists between these levels, cf. the behaviour of affixes such as *-al* (*post-al, 'admir-able ~ ad'mir-able, mort-al*), *-ist* (*typ-ist, scient-ist, dent-ist*), *-ify* (*electr-ify, beauti-fy, ugli-fy*), which operate both on a native and a non-native level. The input to such non-native patterns are actual or potential stems of corresponding lexical items in the sense of (7) above or their English adaptations.

Assuming that stems are a relevant entity in English morphology, if two such stems are combined we get stem-compounds instead of word-compounds, as in (1 a). If they combine with affixes, we get stem-based affixation, as in (1 b, c). There is no reason, therefore, to exclude formations such as *astr-o-naut, mult-i-parous, pol-i-geny, part-i-cide, galvan-o-scope, soci-o-logy, hepat-itis,* etc., from a treatment of English word-formation because of their foreign properties, as Marchand has postulated. Such formations have become increasingly frequent in Modern English and therefore deserve to be regarded as part of the English, though non-native, system, which is stem- and not word-based, but productive in the English language.

4.4. In view of the original situation in Greek, Latin and Neo-Latin, the notion of stem as used here needs some additional clarification, however, since these stems were not adopted without modification. Greek and Latin, like all Indo-European languages, were originally based on a tri-partite morphological structure, root + stem-formative + inflection. In compounding, the root as first element of a compound was followed by a stem-formative, as in *agr-o-nomia*, but, at least in the later stages of Indo-European, could instead also be followed by a composition vowel which goes back to an inflectional ending, as in *agr-i-cultura*. This latter type of formation resulted from lexicalised phrases involving case marking of the first member, which served as the basis of analogical parallel formations, a phenomenon that can also be observed in other branches of Indo-European like Germanic (cf. Kastovsky 2009). In the course of time, however, the stem formatives as well as the original inflectional endings became opaque, in Latin more so than in Greek, and these morphological elements can no longer be identified with any inflectional material but have to be re-interpreted as linking elements functioning in the same way as the German "Fugenelemente" in *Universität-s-bibliothek, Liebe-s-heirat, Kind-er-garten*, etc., which have the same origin. They were imported in the same function into English, where they show up even with native material, cf. *kiss-o-gram, speed-o-meter, Kremlin-o-logy*. Their distribution and the mechanism that triggers them still need systematic investigation.

4.5. It has always been recognised that at least some native prefixes and suffixes have developed from first- and second-members of compounds (cf. *fore-, out-, -dom, -hood, -monger, -wise*), resulting in a synchronic cline between compound constituents, semi-affixes (sometimes called "affixoids") and genuine affixes, cf. Marchand's (1969) analysis of *-like, -monger, -wise* as semi-suffixes. Apparently, the same kind of development happened to first and second members of non-native stem-compounds, whatever their origin, e.g., *neo-, crypto-, multi-* or *-logy, -nomy, -itis*, resulting in the same cline between compounding and affixation as with native word-formation. There is thus no need for undefined terms such as "combining forms" or "terminal elements" for these. Therefore, it is not the

demarcation of combining forms and affixes that is at issue here. The real problem, which, however, has rarely been articulated, is the demarcation of compounding and affixation in general, where a strict dividing line does not seem to exist synchronically. Rather we have to assume a cline both with regard to formal (phonological and/or morphological) and semantic behaviour. For the latter, a more refined definition of "lexically specific" (for words and stems) and lexically non-specific (= general, abstract, etc. for affixes) would be necessary. Perhaps an analysis within the framework of grammaticalisation might be helpful, but whether such a distinction is really viable in view of the fact that affixes can have very specific meanings, cf. *-age* 'fee' in *anchorage, corkage*, and lexemes can have very general meanings like *thing, place* is arguable. In any case, this would be the task of a more systematic word-formation semantics, which is still a desideratum.

4.6. This solution based on the introduction of the notion of "stem" beside the category "word" as possible lexeme representations works for those instances where we can reconstruct a non-native stem as the starting point of the modern formation, as in *astr-o-naut, Mars-naut, Angl-o-phile, audi-o-metry, bibli-o-graphy, dramat-ic, scient-ist,* etc. But not all foreign elements represent such stems in unmodified form; some must be regarded as shortened versions of such stems, i.e., they involve the phenomenon of clipping (cf. Marchand 1969: 441 ff.). Take the case of *Euro-* in *Eurocentric, Eurocrat, Eurocity (train)*, let alone the *Euro* as the currency, where the shortened form has become an independent lexeme, as in *laboratory > lab,* etc. The classical stem is *Europ-(a)*, but this has lost its final consonant in the respective formations. Such shortenings happen with native lexemes as well and lead either to simple clippings or clipping compounds (cf. Marchand 1969: 445 f.), also called stub compounds. Examples are:

(9) *cablegram* (*cable* + (*tele*) *gram*), *paratrooper* (*para* (*chute*) + *trooper*), *Amerindian* (*Amer* (*ican*) + *Indian*), *news stand* (*news* (*paper*) *stand*), *breathalyser* (*breath* + (*an*) *alyser*), *boatel* (*boat* +(*hot*) *el*), *guesstimate* (*guess* + (*es*) *timate*).

Such formations have to be analysed as consisting of one part which is the result of clipping combined with another constituent having word status. There may also be a partial phonological overlap between the constituents as in *boatel, guesstimate*. Such compounds have become relatively frequent in recent decades. Again, we do not have to have recourse to a category of "combining form" for their description, since clipping is a process also affecting simple lexical items such as *bus < omnibus, plane < aeroplane, lab< laboratory*, etc.

There are also instances where both parts are clipped, such as

(10) a. *Chunnel = Ch* (*annel*) + (*t*) *unnel*, *motel = mo* (*tor*) (*ist*) +(*ho*) *tel*,
  b. *Oxbridge* 'the universities of Oxford + Cambridge', *transceiver* 'gadget which is a transmitter + receiver'

Such combinations again are rather heterogeneous. Some reflect the structure of regular compounds. Thus *Chunnel* and *motel* are parallel to determinative compounds, the *Chunnel* being basically a *tunnel* operating underneath the *Channel*, and a *motel* is basically a *hotel* for *motorists*. *Boatel* 'a hotel which is a boat' follows the pattern of copulative compounds like *girlfriend* 'friend who is a girl'. *Oxbridge, transceiver,* on the other hand, can be regarded as being parallel to dvandva compounds, which denote a combination (union) of their respective referents, as in *Austro-Hungary* 'an entity which consists of both Austria and Hungary', *concavo-convex* 'having the property of being both concave and convex', similarly G *taubstumm* 'deaf-mute'. This compound type is an innovation in English, which was borrowed from Latin and Greek (cf. Hatcher 1951), because English had lost this old Indo-European compound type. In OE there are two examples, *aþumswerian* 'son-in-law and father-in-law', *suhtorfædran* 'nephew and uncle', which also have to be interpreted as 'entity which consists of x and y', but otherwise the type was dead.

4.7. The combinations in (10) are often regarded as examples of "blending", defined as "merging parts of words into one new word", e.g., by Marchand (1969: 451), who puts them together with

(11) *smog* = *sm-*(*oke*) + (*f*)*-og*, *brunch* = *br-*(*eakfast*) + (*l*)*-unch*, or Lewis Carroll's *Jabberwocky* words *slithy* (*slimy* + *lithe*), *chortle* (*chuckle* + *snort*), etc.

But there is a difference between (10) and (11); cf. also Hansen *et al.* (1985: 144 ff.). The combinations in (10 a) follow the usual determinant/determinatum schema of determinative or copulative compounds, except that both constituents are clippped, i.e., a *Chunnel* is basically a *tunnel*, etc. The combinations in (10 b) resemble exocentric compounds such as *hunchback, paleface*, where the determinatum is not overtly part of the formation itself and might be interpreted as zero, i.e., *paleface/Ø* = 'someone (= Ø) having a pale face'. Similarly, *Austro-Hungary* can be interpreted as 'entity (= Ø) which consists of both Austria and Hungary', but none of these interpretations applies to (11). These formations usually denote a referent which is a mixture or crossbreed of the referents of the constituent lexemes and therefore constitute real blends from a semantic-cognitive point of view as well. They have no direct counterpart in compounding. Such formations seem to be a relatively recent development, which have extended the range of compounding considerably.

4. Conclusion

To sum up: I have tried to show that the notion of "combining form" introduced by the *OED* is not necessary. The categories of "word", "stem", "affix", "affixoid", "clipping" and "blending" necessary in word-formation for independent reasons are sufficient to deal with the formations in question. Moreover, it can be argued that compounding, affixation, clipping and blending should be regarded as prototypical patterns arranged on a scale of progressively less independent constituents ranging from word via stem, affixoids, affix, curtailed word/stem to splinters as constituents of blends, and finally acronyms (letter combinations), i.e.

(12) compounding (word) > stem compounding (stem) > affixoids > affixation proper (word-/stem-based) > clipping compounds (clipping of words/stems) > blending > splinters > acronyms

This scale interacts with the typological heterogeneity of the English word-formation system, which allows both words and stems as input to word-formation processes. Therefore, the notion "combining form" is something like a red herring in lexicology, because it creates more problems than it solves and should therefore be given up

References

Bauer, Laurie. 1983. *English word-formation.* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Bauer, Laurie. 1998. Is there a class of neoclassical compounds, and if so is it productive? *Linguistics* 36. 403–422.

Bloch, Bernard & George L. Trager. 1942. *Outline of linguistic analysis.* (Special Publications of the Linguistic Society of America). Baltimore: Waverly Press.

Giegerich, Heinz J. 1999. *Lexical strata in English: Morphological causes, phonological effects.* Cambridge: Cambridge University Press.

Hansen, Barbara, Klaus Hansen, Albrecht Neubert & Manfred Schentke, 1985. *Englische Lexikologie. Einführung in die Wortbildung und lexikalische Semantik.* 2., durchgesehene Aufl. Leipzig: VEB Verlag Enzyklopädie.

Hatcher, Anna Granville. 1951. *Modern English word-formation and Neo-Latin.* Baltimore, Md.: John Hopkins Press.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Kastovsky, Dieter. 1986. Problems in the morphological analysis of complex lexical items. *Acta Linguistica Academiae Scientiarum Hungaricae* 36. 93–107.

Kastovsky, Dieter. 1992. Typological reorientation as a result of level interaction: the case of English morphology. In: Kellermann, Günter & Michael D. Morrissey (eds.), *Diachrony within synchrony: Language history and cognition.* (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 14). Frankfurt/M.: Lang, 411–428.

Kastovsky, Dieter. 1994. Historical English word-formation. From a monostratal to a polystratal system. In: Bacchielli, Rolando (ed.), *Historical English word-formation. Papers read at the Sixth National Conference of the History of English*. Urbino: QuattroVenti, 1994. 17–31.

Kastovsky, Dieter. 1996. Verbal derivation in English: A historical survey. Or: Much ado about nothing. In: Britton, Derek (ed.), *English historical linguistics 1994*. (Current Issues in Linguistic Theory 135). Amsterdam/Philadelphia: Benjamins. 93–117.

Kastovsky, Dieter. 2000. Words and word-formation: Morphology in the *OED*. In: Mugglestone, Lynda (ed.), *Lexicography and the* OED. *Pioneers in the untrodden forest*. Oxford: Oxford University Press 2000.110–125.

Kastovsky, Dieter. 2006. Historical morphology from a typological point of view: Examples from English. In: Nevalainen, Terttu (ed.), *Types of variation: Diachronic, dialectal and typological interfaces*. (SLCS 76). Amsterdam, Benjamins. 53–80.

Kastovsky, Dieter. 2009. "Diachronic aspects", in: Lieber, Rochelle – Pavol Štekauer (eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Lehrer, Adrienne. 1998. *Scapes, holics*, and *thons*: The semantics of English combining forms. *American Speech* 73. 3–28.

Lyons, John. 1977. *Semantics*. Vol 1–2. Cambridge: Cambridge University Press.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation*. 2nd rev. ed. München: Beck.

Matthews,Peter H. 1974. *Morphology*. (Cambridge Textbooks in Linguistics). London: Cambridge University Press.

New English Dictionary. 1884–1928. Oxford: Clarendon Press.

Oxford English Dictionary, 2nd ed. 1989. Oxford: Oxford University Press.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Prćić, Tvrtko. 2005. Prefixes vs initial combining forms in English: A lexicographic perspective. *International Journal of Lexicography* 18. 313–334.

Prćić, Tvrtko. 2007. Headhood of suffixes and final combining forms in English word formation. *Acta Linguistica Hungarica* 54. 381–392.

Prćić, Tvrtko. 2008. Suffixes vs. final combining forms in English: a lexicographic perspective. *International Journal of Lexicography* 21 (prepublication)

Shorter Oxford English Dictionary on Historical Principles. 1993. Oxford: Clarendon Press.

Stein, Gabriele. 1973. *English word-formation over two centuries*. Tübingen: Narr.

Stein, Gabriele. 1977. English combining forms. *Linguistica* IX. 140–149.

Warren, Beatrice. 1990. The importance of combining forms. In: Dressler, Wolfgang U., Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary morphology*. Berlin/New York: Mouton de Gruyter. 111–132.

Webster's New International Dictionary. 1994. 2nd ed. Springfield, Mass.

Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)

edited by R. W. McConchie, Alpo Honkapohja, and Jukka Tyrkkö

Cascadilla Proceedings Project   Somerville, MA   2009

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Kastovsky, Dieter. 2009. *Astronaut*, *astrology*, *astrophysics*: About Combining Forms, Classical Compounds and Affixoids. In *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*, ed. R. W. McConchie, Alpo Honkapohja, and Jukka Tyrkkö, 1-13. Somerville, MA: Cascadilla Proceedings Project.

or:

Kastovsky, Dieter. 2009. *Astronaut*, *astrology*, *astrophysics*: About Combining Forms, Classical Compounds and Affixoids. In *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*, ed. R. W. McConchie, Alpo Honkapohja, and Jukka Tyrkkö, 1-13. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2161.