Analogical neologisms in English

Elisa Mattiello

Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
<elisa.mattiello@unipi.it>

The main purpose of this paper is, on the one hand, to show the relevance of the notion of analogy to English word-formation, and, on the other hand, to propose an array of (phonological, morphotactic, and semantic) similarities between analogy-based neologisms and the model words they are based on. The theoretical framework of this paper is Natural Morphology. In particular, we adopt Dressler & Ladányi's (1998: 35) approach to analogy, with a tripartite subdivision distinguishing surface analogy, created on the model of a unique concrete form, from rule productivity, with a precise abstract pattern described in a rule format, and from analogy via schema, with prototype actual words but no exact pattern. Examples of English creative neologisms of the three types – selected from the online collections *Neologisms – New Words in Journalistic Text* (1997-2012) and *The Rice University Neologisms Database* (2004-2014) – are adduced and categorised along scales of similarity between target and model.

The paper discusses analogy in relation to the key notions of creativity, productivity, and rule. It shows that analogy is not devoid of relations to morphological rules, as the oft-cited compound nouns *earwitness* and *whitelist*, respectively after *eyewitness* and *blacklist*, demonstrate. Although analogy is less constrained than rules, it is viewed as a promising area of investigation in word-formation. The paper shows that surface analogies recur throughout the spectrum from rule-based to extra-grammatical formations. It identifies scales of affinities between target and model which allow for the association of the former with the latter. These scales may correspond to different degrees of easiness in both recognising the model (model recoverability) and understanding (or accepting) the target (target disambiguation)*.

KEY WORDS: analogy, similarity scales, neologisms, English

* We would like to address special thanks to Wolfgang U. Dressler for his discussions on the topic of this paper and for his careful reading of an earlier draft. We are also hugely indebted to Marianne Kilani-Schoch and Klaus-Michael Köpcke for their precious comments on the paper and useful bibliographical suggestions. We would also like to thank Antonio Bertacca and Giovanni Iamartino, whose feedback gave us encouragement and advice. Lastly, we wish to extend our thanks to the European Science Foundation (ESF), whose support made this research possible as part of the activity entitled 'European Network on Word Structure (NetWordS): Cross-disciplinary Approaches to Understanding Word Structure in the Languages of Europe'.

 (received October 2015)

1. Introduction

> [A]nalogy is a very powerful tool that is not rare and exceptional but frequently used and that can explain much more than accidental coinages. (Krott 2009: 118)

The notion of analogy pertains to many different disciplines and sub-branches (Lahiri 2000). In grammar, it has commonly been associated with language change (Paul 1880; Hermann 1931; Anttila 1977; Hock 1991; McMahon 1994; Fischer 2007; Aronoff & Fudeman 2011) and grammaticalisation (Traugott & Heine 1991; Hopper & Traugott 2003). More specifically, in inflectional morphology, it has been investigated for its impact on first language acquisition (Clark 2009; Dressler & Laaha 2012), as opposed to grammatical productivity (Dressler & Ladányi 1998; Kilani-Schoch & Dressler 2002, 2005). Recently, the attention of scholars has focused on the relevance of the analogical principle to the sub-module of word-formation, either grammatical (e.g. in compounding, Krott 2009) or extra-grammatical (e.g. in abbreviations and blends, De Smet 2013; Mattiello 2013), or both (Klégr & Čermák 2010). It is to this latter sub-domain that the present paper intends to give its contribution. Indeed, works dealing with analogy in word-formation are still sporadic and not specifically on English. For instance, Krott (2009) is confined to analogical compounds in various languages, including Dutch and German, with a limited number of English examples, and Klégr & Čermák's (2010: 240-241) sample of what they call "presumed analogical formations" includes 344 items drawn from the *Concise Oxford Dictionary* (1995), some of which are, however, "dubious or incongruous" (p. 230) according to the authors. In the present work, the total number of analogies amounts to 488 English examples taken from two online collections (see § 3.2), with only five overlapping items. Furthermore, in a larger project on a related subject with a double number of items (Mattiello, *in preparation*), additional corpora have confirmed the morphological categories, types of analogy, and model-target relationships found in this preliminary study on analogical neologisms.

In word-formation, analogy can be defined as the process whereby a new word is coined that is clearly modelled on an already existing word or on a set of words constituting a word family (i.e. a group of words sharing the same base(s)) or a series (i.e. a group of words sharing the same formation, or a subgroup of the same formation). A new analogical word is commonly called 'target', whereas the source word/set of words it is based on is referred to as either 'model' or 'trigger'. Targets that are obtained via the analogical process typically

share some trait(s) with their models, so that a relationship of partial similarity exists that allows the association of the former with the latter. Targets and models are linked by at least one feature, be it phonological, morphotactic, or semantic.

In particular, this paper focuses on 'surface analogy' ("Oberflächenanalogie" in Motsch 1981: 101), in which similarity is with one precise model word, rather than with prototype actual words (cf. 'analogy via schema' in § 2). This is the type of analogy which occurs in the formation of *software* [1960] (OED2) after the model word *hardware* [1947], or of *white market* [1943], coined after *black market* [1727] to refer to 'authorized dealing in things that are rationed' (OED3, see also *grey market* [1934]).[1] The aim of the paper is, first, to show the role of analogy in English word-formation, and, second, to investigate the mechanisms whereby new analogical forms are coined, and the affinities that they share with their model words. The paper shows that analogical formations are illustrated both by extra-grammatical formations obtained via creative techniques and by grammatical words conforming to productive rules.

In addition, the paper investigates the nature of the relationship occurring between the model and the target, which is usually a relation of partial resemblance. It starts from the assumption that, from the speaker/coiner's perspective, analogy occurs when a similarity is perceived between the constituents of a model and those of a potential target. On the other hand, from the hearer/interpreter's perspective, once an analogical formation is coined, a judgement of partial similarity between the target and the model allows for the association of the former with the latter (cf. "analogical assimilation" in Wanner 2006: 121),[2] and this may facilitate the target interpretation or encourage its acceptability. This can be stated in the following terms:

> T(arget) can be partially associated with M(odel) (T ← M), as long as the two terms share some (or more) trait(s) through which a concrete similarity between them can be established. The greater the similarity, the more likely the model recoverability and the target disambiguation/acceptability.

As far as similarity is concerned, this paper categorises analogies along different scales of similarity, including not only morphotactic affinities (cf. Klégr & Čermák 2010), but also phonological (cf. Kilani-Schoch & Dressler 2002, 2005), and even semantic ones. Therefore, the present categorisation differs from Klégr & Čermák's (2010: 231)

classification, which is purely based on morphotactic distinctions, and detects more subtle differentiations between cases of analogical formations. Indeed, the affinity between the target *oldster* [1818] 'a midshipman who has served for over four years' (OED3) and its model *youngster* [1608], or between *mouse potato* [1994] 'a person who spends large amounts of leisure time using a computer' (OED3) and *couch potato* [1979] is not merely in the derivational or compounding processes obtaining T on the pattern of M (Klégr & Čermák 2010: 231), but also in the phonology and semantics of T and M. For instance, in the case of *mouse potato*, the metaphorical meaning of the model *couch potato* 'a person who spends leisure time passively sitting around' (OED2) is essential for the target interpretation.

We expect that a more refined categorisation based on various parameters can allow for more stable predictions both on the way the analogical process operates (how new words can be formed from the coiner's viewpoint) and on the way the association works (how new words can be related to existing ones from the interpreter's viewpoint). Predictability, indeed, is a discriminatory element distinguishing productivity from analogy, this latter being normally related to lack or, at least, to a lower degree of output predictability (Plag 2003: 38). It is this latter claim that we expect to disconfirm showing recurrences in the creation of analogical neologisms and tendencies that indicate the directions in which analogical processes typically proceed.

In this paper, we assume that the association of the target to its model is particularly favoured by the co-occurrence of similarity at various levels, namely, in phonological and morphosyntactic form, as well as in meaning. Yet, these assumptions require a clearer taxonomy of analogical formations, on which steadier predictions may be founded. More reliable predictions would certainly require quantitative and qualitative analyses of larger databases (cf. § 3) and corpora of attested examples, but quantitative investigations and pragmatic analyses of contextualised data are reserved for a larger project on this subject (Mattiello, *in preparation*). The present project, as section 3 will elucidate, concentrates on existing databases of neologisms.

The paper is organised as follows. Section 2 is a brief overview of the concept of analogy in the relevant literature, especially in the theoretical framework of Natural Morphology. Section 3 describes the data sources and selection, and discusses the notion of neologism vis-à-vis related concepts, such as nonce words and occasionalisms. Section 4 proposes three scales of similarity and progressive levels of association of T with M. We anticipate that similarity scales may

intersect in some analogical cases, whereas other cases may only exhibit similarities restricted to one scale. Section 5 offers a morphological categorisation of the analogical formations analysed in section 4, and then discusses the results of data analysis in terms of regularity of the operation and predictability of new forms. The conclusions we draw in section 6 are meant to justify analogy as a legitimate area of study in word-formation.

2. Theoretical landscape

The notion of analogy originates from the Ancient Greek grammarians Aristophanes of Byzantium and Aristarchus of Samothrace, who adapted its meaning from the mathematical term '*ναλογία* (*analogía* 'analogy') indicating a mathematical proportion. Latin grammarians inherited the Greek term and translated it as *proportio*, *comparatio*, or even as *regula* (Schironi 2007).[3] The nineteenth-century Neogrammarians (e.g. Paul 1880: 106-120) also adopted the Greek proportional descriptive technique, especially used for morphological inflection, according to which A : B = A' : X (X = B').[4]

In the twentieth century, the concept of proportional equation was proposed again by Bloomfield (1933: 275-276), who distinguished "unique analogy" (i.e. analogy with a unique model) both from "irregular analogy" (i.e. with few forms) and from "regular analogy" (i.e. according to a rule). In his view, regular analogies are "habits of substitution" (Bloomfield 1933: 276). Hence, we can analyse *Charlestoner* [1927] 'one who performs the dance called *Charleston*' (OED2) as formed on the regular analogy of *dancer*, *waltzer*, *two-stepper*, and so on. In other words, analogy involves a paradigmatic substitution of the variable part in the proportion *dance*, *waltz*, etc. : *danc(e)-er*, *waltz-er*, etc. = *Charleston* : X (X = *Charleston-er*).

However, in the American generativist tradition of the 1960s-1970s (especially Kiparsky 1974; Aronoff 1976), analogy was considered an illegitimate topic in linguistics and replaced by more adequate notions which allowed for a high degree of generalisation. Nevertheless, the neogrammarian notion of analogical formation ("Analogiebildung" in Paul 1880) had not disappeared and it came back as a legitimate area of enquiry. Charles Hockett, in particular, was the first to defend Bloomfield's (1933) concept of analogy: "An individual's language, at a given moment, is a set of habits – that is, of analogies" (Hockett 1968: 93). Another unequivocal defender of analogy was Anttila (1977).

The end of the twentieth century and the beginning of the current century are characterised by two opposite tendencies dealing with analogy. On the one hand, authors of standard descriptions of present-day English word-formation, such as Bauer (1983), Plag (2003), and Bauer *et al.* (2013) mention only in passing the role of analogy in vocabulary expansion in English, especially relating it to back-formation. However, Bauer *et al.* (2013: 518-530) also discuss analogy in a chapter devoted to paradigmatic processes, as part of what they call "paradigmatic morphology" (p. 519).

In general, scholars tend to minimise the importance of analogy by associating it to creativity and unpredictability, rather than to rule-governed productivity (see, e.g., Bybee 2010: 57). As Ladányi (2000: 2) observes, the foundation of this view is that analogy is often conceived as a surface means to produce occasionalisms (see Christofidou's 1994 "Okkasionalismen") or hapax legomena (i.e. "words with only a single attestation", Bauer 1988: 65) via particular defaults rather than productive rules.

The opposite tendency regards analogy as an essential concept in word-formation. For instance, Zemskaja (1992), as a prominent example of the Russian tradition of word-formation, claims that analogy is the most important means of derivation including both its productive and unproductive forms. In her opinion, the only difference between the two is that, in the case of productive derivation, analogy works via rule types, while in the case of unproductive derivation it usually works via defaults of individual (complex) words. Blevins & Blevins (2009: 10) generalise the same view by claiming that, from a traditional perspective, "a rule can be understood as a highly general analogy". This is also the position of Krott (2009: 118), as the above-mentioned quote suggests. Van Marle (1990: 267) even tries to reconcile the three concepts of analogy, creativity, and rule by mentioning the ability of "rule-creating creativity", which, according to him, directly bears upon analogy as a synchronic morphological force. Zemskaja's (1992) view about the general role of analogy in derivation as well as van Marle's (1990) stress on the importance of the analysis of existing words for the study of morphological creativity may be taken as motivations to reconsider traditional dichotomies, such as analogy *vs* rule and creativity *vs* productivity.

In this paper, it is claimed that surface analogy differs from productive rules in several respects. First, surface analogy is based on concrete models of precise similar forms (as in Fr. *alun-ir* 'to land on the moon', *amerr-ir* 'to land on the sea', formed on *atterr-ir* 'to land', cited in Kilani-Schoch & Dressler 2002: 298), rather than on abstract

patterns (Fertig 2013). Second, analogy is output-orientated, whereas rules are also input-orientated (Dressler & Laaha 2012: 49). Third, in an analogical approach to word-formation there is no concept of 'potential word' (Aronoff 1983), because all words are potentially created, or possible, as long as there is an actual or existing word which can function as model. By contrast, words derived through a productive rule are constrained by precise restrictions on both the base and the output (see Aronoff's 1976 'unitary base constraint' and 'word-based word-formation', or Scalise's 1988 'unitary output constraint').

Our position is close to Bauer's (2001) view that the two notions of rule and analogy coexist within word-formation, in that neither can underlie morphological innovation by itself (cf. Plag 1999). Although analogy may fail to make suitable predictions and, unlike rules, is more permissive in terms of input and output categories and structure, it certainly plays a considerable role in both language change (see McMahon 1994) and word-formation (see Szymanek 2005: 431; Klégr & Čermák 2010: 230; Miller 2014: 88-89). In particular, whereas rules can only explain the functioning of productive (or no longer productive) morphological processes (see productive *-er* in *speak-er*, or unproductive *-th* in *warm-th*), analogy can motivate both grammatical word-formation and the mechanisms involved in creative formations, because it can be viewed as the reason for a superficial similarity between two structures.

By contrast, computational analogical models, such as Skousen (1989, 2009) and Becker (1990), do not contemplate the notion of productivity and equate classical (generative) morphological rules with the notion of analogy. In the present paper, instead, we distinguish grammatical (rule-governed) surface analogies of the type *whitelist* [1842] 'a list of people or things considered acceptable' (OED3) – obtained after *blacklist* [1624], but also conforming to rules – from extra-grammatical ones (e.g. *yettie* [2000] 'young, entrepreneurial, and technology-based [person]' (The Guardian) formed after the acronym *yuppie*), which are instead obtained via extra-grammatical operations.

The theoretical framework best suited for describing the linguistic status of most analogical word-forms is extra-grammatical morphology. The latter is a cover term provided by Natural Morphology (Dressler 2000) that applies to a set of heterogeneous formations "which do not belong to morphological grammar, in that the processes through which they are obtained are not clearly identifiable and their input does not allow a prediction of a regular output" (Mattiello 2013: 1). Acronyms, as well as blends, initialisms, clippings, hypocoristics, reduplicatives, back-formations, and expletive infixes are widely

accepted examples of extra-grammatical morphological phenomena.

Word creation is another concept that is relevant to the production of analogical words. Ronneberger-Sibold (2000, 2008), for instance, has opposed regular (productive) word-formation to word creation, i.e. the production of new lexemes via intentional extra-grammatical operations, such as shortening, extra-grammatical derivation, and blending, in domains such as humorous literary texts and brand names. However, not all analogical words, as we will see, are obtained through extra-grammatical "creative techniques" (see Ronneberger-Sibold 2008: 203-205 for a distinction between creative techniques and regular rules or models of word-formation). Other analogies follow a schema identifiable as a series of concrete words exhibiting the same formation, or as a word family with identical bases.

Therefore, in this paper Bauer's (1983: 96) definition of "genuine" analogical formation (comparable to 'surface analogy') as one which does not give rise to productive series is considered as too narrow to include the whole range of English neologisms whose origin can be ascribed to analogy. Actually, analogy often gives rise to productive series, as with the bound morpheme *-licious* (often preceded by vowel), which was originally a 'splinter' (Lehrer 1996, 2003) from *delicious* used in new blends, but has recently been assigned the label of 'combining form' in OED3. Indeed, its use has become regular in the formation of attested neologisms with the sense 'embodying the qualities denoted by the first element to an attractive degree', as in *babelicious* [1991], *bootylicious* [1994] 'sexually attractive' (OED3), *groovalicious* [2002] (OED3), Lehrer's (2007) jocular *blendalicious*, and the newly coined *Hooterlicious* [2008] (used for 'good-looking women at *Hooters*' by Rice University students). These words constitute a productive series which has originated a schema. Hence, unlike surface analogies, new words coined after this schema have not a unique model.

As for schemas, a theoretical approach that deals with the relation between analogy and schemas is Booij (2010), who develops a model known under the name of Construction Morphology. In Booij (2010), it is claimed that schemas and subschemas may operate on symbolic features, and that the crucial difference between analogical formations and schema-based formations lies in their making reference to different degrees of abstraction. Analogy in this model is defined as a strictly local mechanism, which is complementary to schemas and may constitute an initial stage of the development of a schema (Booij 2010: 88-93). From this perspective, surface analogy

may be viewed as the trigger of a schema, in that all series originally owe their foundation to a unique model, which later produces similar formations and develops into a schema.

The role of diachronic language development is therefore crucial for the model of analogy adopted in the present paper. For instance, the above-mentioned case of *-licious* words deserves special attention from the diachronic viewpoint. In particular, it shows how a word end (i.e. a splinter) merging with other words (or word parts) to form new blends has become frequent in use, regular, productive, and has therefore developed into a final combining form. Fradin (2000: 37) would even include *-licious* in the category of "secreted affixes" (e.g. *-gate* or *-holic*), which exhibit a certain level of abstraction and allow generalisations. However, as will be explained below, new *-licious* words are not coined after an abstract pattern, but still depend on concrete forms. In Natural Morphology, combining forms, such as *-licious*, *-gate*, etc., can be accommodated within that part of "marginal morphology" (Dressler 2000: 7) concerning the internal boundaries, in particular, among phenomena that are transitional between derivation and compounding.

Another concept which needs to be discussed in this theoretical section is "reinterpretation" (Hock 1991: 176) or "reanalysis" (Hopper & Traugott 2003: 56; Booij 2005: 262). Indeed, some analogical formations are often preceded or triggered by morphological reanalysis. For instance, in the oft-quoted example of *cheeseburger* [1938] (OED2), created after *hamburger* [1889] (Plag 2003: 37), proportion has been preceded by re-segmentation. Indeed, *hamburger* was originally coined from the name of the German city of Hamburg, for 'a native or inhabitant of Hamburg' (OED2). Yet, the derived word (with a different meaning) has later been reanalysed as *ham* + *burger*, thus giving rise to the series which includes *beefburger* [1940], *chickenburger* [1936], *eggburger* [1960], *vegeburger* [1945], and the above-mentioned *cheeseburger*. Reanalysis is also in the verb *boycott* [1880] (OED3), originally from the name of Captain Charles C. Boycott, but later reinterpreted as *boy* + *cott* before the analogy *girlcott* [1884] 'of a group of women: to boycott' (OED3) was formed by feminists. Broadly speaking, also the reinterpretation of *delicious* as *de-* + *-licious* has given birth to the corresponding series of *-licious* words. Indeed, according to the OED, †*licious* [*c*1420] was an obsolete aphetic form of *delicious*. However, not all analogies involve reanalysis: e.g. Plag's (2003: 37) proportion *sea* : *sea-sick* [*a*1566] = *air* : X, X = *air-sick* [1785] (OED3) does not. Nor all combining forms involve reanalysis and identification with existing morphemes: e.g. *-holic*, from *alcoholic*,

does not correspond to any existing English morpheme; hence, *alcoholic* has not been reanalysed as *alco-* + *-holic* in the formation of the latter combining form.

Lastly, we disagree with Bauer (1983: 96) and Krott (2009: 118), who consider formations such as *ambisextrous* [1929] 'ambisexual' (OED2) and *wargasm* [n.d.], as analogy coinages because of chance phonetic resemblance with single exemplars (i.e. *ambidextrous*, *orgasm*). Actually, these are blends merging two words where they overlap: in *ambisextrous* the shorter word *sex* is intercalated within the longer one (*ambidextrous*, entirely present in the blend except for the consonant *d*), whereas in *wargasm* the two constituents *war* and *orgasm* share the same vowel sound /ɔː/ making them combine as one word. Therefore, while from the viewpoint of phonological similarity, the blends resemble (and evoke) their longer constituents, from the viewpoint of morphotactic similarity the analogy is with other blends of the same intercalative and/or overlapping type (e.g. *entreporneur* [n.d.] 'entrepreneur + porn', Mattiello 2013: 120, *slanguage* [1879] 'slang + language' OED2). It is clear that a blend like *chillax* [2008] 'to relax or take a break' (from *The Rice University Neologisms Database*, henceforth **RUND**, 2004-2014) is phonologically similar to (quasi-homophone with) *relax*, but this is not because the former is constructed by analogy with the latter, but rather because the latter, *relax*, combines with *chill* to obtain the former, *chillax*.

Hence, the above-mentioned types of blend, *ambisextrous* and *wargasm*, share with their models *entreporneur* and *slanguage* the same morphotactic structure involving overlapping constituents. In other words, they are "overlapping blends" (Mattiello 2013: 122), i.e. blends which exhibit a phonological overlap of vowels, consonants, or syllables between the constituents, with or without a proper shortening. In *wargasm*, for instance, neither *war* nor *orgasm* have been shortened in the new formation, at least from the phonological viewpoint. Indeed, one of the criteria of well-formedness for blends is "recoverability", that is, they must preserve as many segments from the source words as possible (Mattiello 2013: 140), as it happens with *wargasm* and its model *slanguage*, from *slang* and *language*.

This preliminary investigation on the notion of analogy in the creation of new morphologically complex or derived words seems to confirm van Marle's (1990: 268) claim that "analogy, even within the realm of derivational morphology, is no homogenous concept". More precisely, we accommodate the multi-faceted concept of analogy within the tripartite model elaborated by Dressler & Ladányi (1998), partially after Motsch (1981), and distinguish among:

(1) SURFACE ANALOGY, i.e. analogy formed on the pattern of a precise actual word and word form (Dressler 2003: 32; Dressler & Laaha 2012: 49), as in the derived verb *prepone* [1913] 'to bring forward to an earlier time or date' (OED3), after the model word *postpone* [1496], and the blend noun *boatel* [1956] 'boat + hotel' (OED2), after *motel* [1925].

(2) ANALOGY VIA SCHEMA (cf. Köpcke 1993), i.e. analogy formed on prototypes identifiable as actual words and belonging to a schema (i.e. a series or a word family, cf. Wittgenstein's notion of 'family') but no abstract pattern.[5] This category can be further subdivided into:

(a) unstable schemas that are not attested in dictionaries, as with splinters frequently occurring in blends (e.g. *docu-* ← *documentary*, in *docudrama* [1961], *docusoap* [1979], and *net-* ← *Internet*, in *netizen* [1984], *netiquette* [1982]), and

(b) more stable schemas, as in the above-mentioned final combining forms *-burger* and *-licious*, or in *-tainment* (← *entertainment*) found in *edutainment* [1983], *infotainment* [1980], etc.

The sub-sets in (a) and (b) constitute a true continuum, with possible shifts from one sub-category to the other depending on diachronic reasons. For instance, we can predict a diachronic development from a frequent splinter, such as *-ercise* ← *exercise*, to a combining form. Indeed, this element has become productive in English, so much so that Baldi & Dawar (2000: 968) have assigned it the label of "unconventional suffix" used, for example, in *boxercise* [1985] and *dancercise* [1967], after *sexercise* [1942] (see also *deskercise* [n.d.] found in Lehrer 2007: 117; cf. § 4.2). The diachronic criterion, therefore, helps establish target and model in surface analogy and the model word(s) for a series in analogy via schema.

Other scholars discuss the degree of abstraction of such combining forms as *-gate* (from *Watergate*) and *-holic* (from *alcoholic*), which have a stable meaning and a fixed phonological representation, and, therefore, may be considered to be affixes or affix-like constituents (Fradin 2000: 37). Actually, although generalisations are possible for these constituents – e.g. *-gate* denotes 'a scandal involving a cover-up X' and *-holic* 'a person addicted to X' – the level of abstraction that we have in these formatives is not the same as we have in affixes. In other words, they create productive series, but their models are still concrete forms, namely, *Dallasgate* [1975], *Koreagate* [1976], *Hollywoodgate* [1978], etc. for the former and *workaholic* [1947], *milkaholic* [1955], *sugar-holic* [1955], etc. for the latter.

It is also worth noting that series are often time-restricted and may be group-restricted as well. In other words, when a series starts to exist, its expansion is generally rapid at the beginning, especially within specific groups, and may gradually become slower as time goes by. The combining form *-licious*, for instance, has become productive in the twenty-first century, especially among young speakers, whereas *-gate* was much more productive in the newspapers of the 1970s-1980s, to denote scandals comparable with the Watergate scandal of 1972. For this reason, the number of items included in the database of this study has not obstructed the identification of schemas. Furthermore, the homogeneity of the two collections in terms of 1) informants for RUND (see § 3.2), and 2) textual genre for *Neologisms – New Words in Journalistic Text* (henceforth **NEWJT**, 1997-2012) (see § 3.2) has facilitated the detection of similar formations identifiable as series or word families.

(3) Rule productivity or "productivity of abstract patterns" (Gardani 2013: 18). With rule productivity, there is no need of an actual model, in that the rule-governed creation of a new word depends on the precise abstract pattern or template (describable in a rule format). For instance, in *soapery* [1674], which is morphotactically and semantically comparable to *bakery* [*c*1820], the abstract pattern of *-ery* denominal derivation applies (see Marchand 1969: 282-285; Lieber 2005: 385). Indeed, diachronically, *bakery* cannot be the model for *soapery*.
Productivity, therefore, must be kept distinct from creativity. The former is that property of language which allows a native speaker to create new words in a rule-governed way. The latter, on the other hand, is the native speaker's ability to extend the language system in a motivated, but unpredictable (non-rule-governed) way. Hence, productivity is commonly defined in terms of type frequency, semantic coherence, and the property of a process to be used to coin new complex words (Bauer 2001; Plag 2003). Creativity, instead, can be negatively defined as lack of predictability, as with illegal formations (Bauer 2001). However, creative analogical formations can be partially predicted by means of similarity relationships with their models.

In this paper the focus is on analogies of the first type (and partially on analogies of the second type), although (2) can be viewed as an extension of (1), from a single model to a group of prototype words that share the same model. Moreover, while cases of rule productivity are not considered, the possibility of co-existence of surface analogy and rule is admitted ((1) and (3)). As observed by Dressler & Laaha (2012: 49) "surface analogies are not devoid of relations to morphological rules". Thus, *earwitness* [1539] was formed analogically after English

eyewitness [1539] for 'a person who testifies to something on the evidence of his own hearing' (OED3),[6] but it was also formed according to the rule of subordinate noun compounding. By contrast, after *welfare* [1357], a surface-analogically antonym *illfare* [*a*1425] was coined, with different rule patterns in the model and the target: namely, [*well*]$_{Adv}$ + [*fare* 'voyage']$_{N}$ (OED3) *vs* [*ill*]$_{A}$ + [*fare* 'condition']$_{N}$ (OED2) (Kilani-Schoch & Dressler 2005: 133).[7]

3. Methodology

3.1. The concept of neologism

The selection of appropriate data is the most challenging task in the case of analogical neologisms. First, because analogy represents a process still open to much debate, so, what may be considered an analogical formation by some scholars may not by others (see § 2). Second, because neologisms, or new words, are often mere cases of occasionalisms, i.e. words that are coined for single occasions and fail to enter ordinary language. Thus, it is up to the lexicographer to discriminate between words which deserve an actual entry in dictionaries and those which do not.

Dressler (1993: 5028) reserves the term "neologism" for "new words which are meant to enrich the lexical stock of a language (or which are already accepted as such)", as distinct from "occasionalisms" or "nonce words" referring to *ad hoc* formations that are not recognised by the speech community of a language. Occasionalisms, in fact, appear only once and are mainly produced for specific textual effects. The function of neologisms, instead, is the enrichment of a language lexicon. From this viewpoint, it is debatable whether or not extra-grammatical formations, such as clippings or acronyms, can be viewed as true neologisms, in that they generally provide more informal or specialised variants of existing words, rather than new words. By contrast, blends provide new words for novel concepts, so their status as neologisms is less controversial. In general, neologisms can be identified in corpora through their representativeness and token frequency. Type/token frequency are relevant concepts to analogy investigation in corpus linguistics analysis (see Mattiello, *in preparation*, for type/token frequency of new analogical words in corpora of English, such as *British National Corpus* (**BNC**), *Corpus of Contemporary American English* (**COCA**), and others).

Here the term neologism is used in a less narrow sense, and also includes words which, although not being part of everybody's ordinary

language yet, may have occurred in newspapers, films, books, and the media at large, or may be used by restricted communities of speakers, such as Rice students, in their everyday conversations. The typical areas of occurrence of neologisms which we have identified are:

- Literary (esp. poetic) language: novelists and poets are allowed to invent audacious neologisms which are often mere nonce words with an aesthetic function, but may also become part of the literary heritage of a people.[8]

- Specialised language: scientists, physicians, linguists, and economists coin new words to name new discoveries or illnesses, recently formed research teams or corpora, modern phenomena, and tendencies. Since the concepts are new, a novel terminology is needed to refer to them. These novel words often become part of a specialised jargon that is used or understood only by a restricted speech community sharing common profession or occupation.

- Journalistic/Advertising language: journalists and advertisers coin new words to attract their audiences, to raise their interest, and have a strong impact on their memory. Journalistic and advert neologisms are often occasionalisms with a jocular or playful effect.

- Young people's language: young people and students tend to create new words for communicating with their peers without being understood by their parents, professors, or adults in general. They mainly use creative neologisms to express intimacy with their group, to exclude outsiders, or simply to show off.

In this paper, only the latter two areas will be explored. English neologisms are also distinguished diachronically into:

- Past neologisms: In the early modern period (late 15th c.-late 17th c.), there was a massive vocabulary expansion in English, with neologisers who contributed to introduce new words into the English lexicon coming from Latin (e.g. *data* [1645], *hostile* [1597], *popular* [1589], *stimulus* [1684]) or French (e.g. *civilisation* [1656], *elegant* [*c*1475], *regime* [*c*1475], *role* [1606]). Words such as *popular* or *elegant* represent neologisms only from a historical viewpoint, in that they are not felt to be new nowadays, although their etymology can clarify the word source and history.

- Recent neologisms: In the last two or three decades, new technologies and inventions, as well as the media have contributed to further enrich the English language with newer words, such as *acid jazz*

[1988] 'a genre of dance music' (OED3), *blog* [1999] 'weblog' (OED3), *e-reader* [1995] 'a hand-held electronic device used for reading e-books' (OED3), *SMS* [1991] 'Short message service' (OED3), etc. These words have later been recognised as part of the institutionalised language and therefore codified by dictionaries.

- Present neologisms: Presently, the growth of the English vocabulary is still underway, with analogy playing a fundamental role in the coinage of the latest words which crop up in the news, or in TV shows, sit-coms, blogs, social network sites, or everyday conversation. Some currently coined words may be classified as nonce words or occasionalisms in the traditional sense, because they have expressly been coined for a single occasion and tend to vanish as rapidly as they have been created. For instance, the new compound adjective *blank-American* [2008] 'a white American girl', after *African-American*, is rather ephemeral and attested only in the Rice collection (see RUND in § 3.2). Other new words, however, are given a chance to become true neologisms by being re-used by speakers, and still others will probably have a locus in dictionaries in the near future. The noun *advertainment* belongs here: although it is not attested in the OED, it occurs six times in The Guardian archive. So, like its analogous formations *docutainment* [1978] (OED2), *edutainment* [1983] (OED2), and *infotainment* [1980] (OED3), it is likely to become an institutionalised neologism (see "institutionalization" in Brinton & Traugott 2005: 45).

3.2. Data selection

The database used in this study consists of both recent and present neologisms, since they are both thought to be relevant to an investigation on morphological analogy. It has been compiled selecting analogical words from two existent collections of neologisms available online, i.e. NEWJT and RUND.

NEWJT is a collection of 819 neologisms selected from two newspapers, The Independent and The Guardian, and catalogued by year at Birmingham City University by the Research and Development Unit for English Studies (RDUES) coordinated by Matt Gee, Andrew Kehoe, and Antoinette Renouf. As specified in the RDUES web page, the words have been identified as being new by software developed by the Unit during the APRIL (Analysis and PRediction of Innovation in the Lexicon) project.[9]

In NEWJT, the words are given together with the news extracts in which they have been used. Although the exact source is not indicated each time, it is specified that words collected before 2000 have been drawn from The Independent, whereas those after 2000 have

been taken from The Guardian. The compilers have not provided information about the etymology of the words or the word-formation processes forming them. However, it was often possible to determine the model of surface analogy because model and target tended to co-occur in the same micro-context (sentence).

In NEWJT, word selection was made manually, cross-checking each neologism in the OED and verifying its possible attestations, etymology, and uses. Analogical words are generally indicated as being created 'after the word X' in the OED etymology. Analogies via schemas generally exhibit an attested combining form.

The manual selection has produced the following quantitative results:

- The database totals 95 analogies out of 819 items. In other words, 11.59% of neologisms are analogy-based. Among them:

- 15.78% (15 occurrences) are pure surface analogies (type (1), § 2) involving reanalysis of the model or extra-grammatical processes. This datum confirms the anticipation of the pertinence of extra-grammatical morphology and word creation to surface analogy (§ 2);

- 17.89% (17) are surface analogies which also conform to rule patterns (types (1) and (3), § 2). This datum corroborates the hypothesis that some new words are primarily motivated by the superficial resemblance with a unique model word, but can also be simultaneously motivated by a rule pattern.
The former and the latter sub-types of surface analogy include an 8.42 percentage of recent neologisms that are also attested in the OED.

- 66.31% (63) are, instead, analogies via schema, with 4 obeying unstable schemas (type (2a), § 2), and the remaining 59 items exhibiting attested combining forms (type (2b), § 2). Combining forms belong to marginal morphology, especially to the transition between derivation and compounding (see "marginal morphology" in Dressler 2000: 6-7), and show how the development of established series is often the result of type/token frequency (see, e.g., the diachronic development of *-licious* from a splinter to a combining form or secreted affix, Fradin 2000).

RUND, with currently 9,016 entries,[10] is Suzanne Kemmer's dictionary of neologisms collected over the years by English Linguistics students at Rice University. Only Rice University students can add or delete their entries by logging in. Thus, all compilers must have a val-

id Rice NetID, which excludes the general public from the database compilation process. However, since all compilers add their terms independently, some of the entries coincide: e.g. the adjective *ginormous* is one of the most frequent occurrences in RUND, accounting for 32 entries in the word list, and different spellings (e.g. *bootylicious vs Bootylicious*) account for two separate entries as well. Therefore, in order to have a clearer idea of the actual number of items making up the database, we have downloaded the whole list and accurately cancelled equivalent items. The real number of items in the Rice database has then decreased to 6,755.

For each entry, compilers have provided information about the word's part of speech, the morphological or semantic process obtaining it, its description, its etymology, its use in context, and the source from which it has been taken. Sources are various and range from spontaneous conversation to book titles, magazines, comics, web pages, e-mails, Facebook, TV shows, and the like.

Since RUND is a larger database than NEWJT, for the identification of analogies the selected word list was initially restricted via an advanced search, selecting 'analogy' as word-formation type. This initial selection was then followed by a manual selection discriminating between morphological analogies and semantic ones (e.g. metaphorical extensions). This procedure, however, left out of the selected data many words which had not been labelled by the compilers as formed via 'analogy', but which actually were either surface analogy or analogy via schema. Close reading of all 6,755 selected entries was therefore essential for a complete data collection.

The ultimate selection has produced the following results:

- The database totals 398 analogies (5.89%) out of 6,755 entries. Among analogy-based terms:

- 34.42% (137 entries) are pure surface analogies (type (1), § 2) obtained via extra-grammatical operations or reanalysis;

- 16.58% (66) are surface analogies which also comply with rules (types (1) and (3), § 2). Here the percentage of regular recent neologisms attested in the OED is higher (27.69%) than the 4.41 percentage of attested pure surface analogies;
Remarkably, the sub-type of pure surface analogy in this collection doubles the amount of the sub-type conforming to rules. Creative extra-grammatical morphology proves in this way its suitability to analogical formation.

- 48.99% (195) are analogies via schema, with 12.88% attested words (recent neologisms), and 73.84% exhibiting well-established combining forms (type (2b), § 2).

In the manual selection, however, some cases remained ambiguous between surface analogy and analogy via schema. For instance, the blends *Wasian* [2008] 'white + Asian', *Chrasian* [2013] 'Christian + Asian', and *Fasian* [2013] 'fake + Asian', all modelled on *Blasian* [2008] 'black + Asian' and initially considered surface analogies, share the second constituent and seem to have originated a series. However, the actual use of these words out of the Rice community may be an argument against their being classified as analogies via schema. Schemas indeed imply regularity and generalisation.

Interestingly, in this distribution, the prevalence of surface analogies obtained through extra-grammatical operations over those that also conform to rule patterns suggests that the type of analogy with a precise model is particularly prolific within extra-grammatical morphology (Mattiello 2013). Analogy via schema, by contrast, is best and more frequently illustrated by combining forms, which are part of marginal morphology.

For all RUND words we have also checked attestation in the OED and found out that many of them are mere humorous occasionalisms used only once, or that they belong to a restricted speech community, or even to some student's idiolect. Hence, for some of the words, we have also checked token frequency in The Guardian and The Independent archives, in order to establish the actual level of representativeness that the words have.[11] Apropos, it is interesting to note that a word such as *beefcake* [1949] '(a display of) sturdy masculine physique', which both OED2 and RUND consider a neologism on the pattern of slang *cheesecake* [1934] 'display of the female form' (OED2), now records 481 occurrences in The Guardian and 82 in The Independent. On the other hand, the word *Brangelina* [2008/2013], a blend from 'Brad (Pitt)' and 'Angelina (Jolie)' created by analogy with *Bennifer* [2008] 'Ben (Affleck) + Jennifer (Lopez)' (RUND, see § 4.2), is not attested in the OED, although it occurs 589 times in The Guardian and 88 times in The Independent. An actual attestation of neologisms in dictionaries is, for the synchronic study of analogy, not strictly relevant, in that it is evident that a word such as *Brangelina* has a recognised status in journalistic (even ordinary) language and is not a nonce term. Furthermore, together with *Bennifer*, it has set the pattern for the recently coined nickname *Merkozy* [2011], which has been produced, by analogy, to refer

to '(Angela) Merkel' and '(Nicolas) Sarkozy' (NEWJT-RUND). In the present paper, cases of words such as *Brangelina* and *Merkozy* will be considered present neologisms and classified among morphotactic and semantic analogies (§ 4.2-4.3).

4. Similarity scales

This section aims to provide a taxonomy of possible analogies in our database and in word-formation in general. As stated already, surface analogy involves a relationship of partial similarity between a newly coined word and an actual word or word form which serves as model. Analysing our data, we have observed three different scales of similarity linking the new word to its model. The first one is phonological, and accounts for the association of *symphomaniac* [2008] 'a person that feels extreme devotion and appreciation for the orchestra' (RUND) with the established word *nymphomaniac* [1867] (OED3), differing only for the initial consonant sound. The second one is morphotactic, and accounts for the association of *VTD* [2008], an initialism for 'verbally transmitted disease' (RUND), with the analogous *STD* [1974] 'sexually transmitted disease' (OED2). Here there is not only phonological similarity allowing the association, but also sameness in the extra-grammatical process (i.e. acronym formation) obtaining T after M. The third similarity scale along which analogies can be classified concerns semantic resemblance. For instance, it is by semantic similarity that we associate the neologism *daughterboard* [1971] 'a printed circuit board on which are mounted some of the subsidiary components of a microcomputer' (OED2) with its model word *motherboard* [1965], although the two words also share the same compound head *board* (morphotactic resemblance) and, therefore, the final unstressed syllable (phonological/prosodic resemblance). In the differentiation of the three levels of analysis, only one level at a time will be taken into consideration in § 4.1-4.3, reserving the treatment of cases involving more levels to § 4.4. Between competing levels of analysis, the most prototypical/representative case for each category will be shown. Syntactic similarity (same part of speech) has not been taken into consideration since it seems to be a prerequisite for two forms to be considered analogous.

4.1. Phonological similarity

On a scale of phonological similarity, we can distinguish progressive degrees of resemblance between T and M. The following catego-

ries and sub-categories illustrate a gradual model from the highest to the lowest degree of similarity according to which we can accommodate the analogies of our database.[12]

(1) HOMOPHONE WORD: the target is a homophone (but not a homograph) of the model in *buysexual* /baɪˈsek.sjuː.əl/ [2013] 'a type of person who is supposedly attracted or will perform sexual acts with people who buy them things' (RUND), from *bisexual* /baɪˈsek.sjuː.əl/ [1914] (OED2);

(2) QUASI-HOMOPHONE WORD (OFTEN WITH A SHARED MONOSYLLABIC HEAD): the target is a quasi-homophone word of the model. T and M can differ for:

(a) a short vowel in an unstressed syllable, as in *Internot* /ˈɪn.tə.nɒt/ [2008] 'a person who refuses to use the Internet' (RUND) ← *Internet* /ˈɪn.tə.net/ [1974] (OED3);

(b) a long vowel or a diphthong in an unstressed syllable, as in *barsexual* /bɑːˈseks.juː.əl/ [2013] 'a heterosexual female who will make-out with someone of the same sex at bars or parties in order to gain attention' (RUND) ← *bisexual* /baɪˈseks.juː.əl/ [1914] (OED2);

(c) a long vowel in a stressed syllable, as in *wordrobe* /ˈwɜːd.rəʊb/ [2008] 'a person's vocabulary' (RUND) ← *wardrobe* /ˈwɔːd.rəʊb/ [?*a*1400] (OED2);

(d) a consonant, as in *rockumentary* /ˌrɒk.jʊˈmen.tᵊr.ɪ/ [1969] 'a documentary on the subject of rock music' (OED3, [2008] in RUND) ← *mockumentary* /ˌmɒk.jʊˈmen.tᵊr.ɪ/ [1965] 'a film, television programme, etc., which adopts the form of a serious documentary in order to satirize its subject' (OED3), with an initial trill, rather than nasal sound, and *Mexploitation* /ˌmeks.plɔɪˈteɪʃn/ [2013] 'a movie that uses Mexican concepts in an exploitation film' (RUND) ← *sexploitation* /ˌseks.plɔɪˈteɪʃn/ [1924] (OED3), with an initial nasal, rather than fricative sound. Also *typoglycemia* /ˌtaɪ.pəu.glaɪˈsiː.mɪə/ [2008] 'the ability to recognize and understand typed, nonsensical, misspelled gibberish' (RUND) ← *hypoglycemia* /ˌhaɪ.pəu.glaɪˈsiː.mɪə/ [1894] (OED2), and the above-mentioned *symphomaniac* /ˌsɪmp.fəuˈmeɪ.nɪ.æk/ ← *nymphomaniac* /ˌnɪmp.fəuˈmeɪ.nɪ.æk/ belong here, with different initial consonants in the target and model. In *tweetheart* /ˈtwiːt.hɑːt/ [2010] 'someone who loves Twitter and uses it a lot' (RUND) ← *sweetheart* /ˈswiːt.hɑːt/ [*c*1290] (OED2), a plosive replaces a fricative in the initial cluster and there is an orthographic adaptation. By contrast, in *tweenager*

/ˈtwiːnˌeɪ.dʒə/ [1949] 'a child who is nearly, or has only just become, a teenager' (OED3, [2000] in NEWJT) ← *teenager* /ˈtiːnˌeɪ.dʒə/ [1941] (OED2), there is no consonant substitution but the addition of a semi-consonant. In *Failbook* /ˈfeɪl.bʊk/ [2013] 'derogatory expression for Facebook' (RUND) ← *Facebook* /ˈfeɪs.bʊk/ [2004] (Wikipedia) and *ego-tourism* /ˈiː.gəʊˌtʊə.rɪ.zm/ [1997] 'tourism around the self' (NEWJT) ← *ecotourism* /ˈiː.kəʊˌtʊə.rɪ.zm/ [1982] (OED3), the consonant change is non-initial;

(e) a consonant cluster, as in *trit* /trɪt/ [2008] 'a unit for measuring information' (RUND) ← *bit* /bɪt/ [1948] 'binary digit' (OED2).

(3) Quasi-rhyming word and same disyllabic head (with same second and third syllables): the target is a compound whose disyllabic head corresponds to the model's head, whereas the modifiers are quasi-rhyming words, as in *blamestorming* /ˈbleɪmˌstɔːm.ɪŋ/ [1997] 'the process of investigating the reasons for a failure and of apportioning blame' (OED3, [2008] in RUND) ← *brainstorming* /ˈbreɪnˌstɔːm.ɪŋ/ [1907] (OED3).

(4) Polysyllabic word with first rhyming syllable and same third syllable: the target is a neoclassical formation whose combining form corresponds to the model's neoclassical combining form and which also shares rhyme with the first syllable of the model word, as in *Hellograph* /ˈhel.əʊ.græf/ [1998] 'an appellative for the Daily Telegraph' (NEWJT) ← *Telegraph* /ˈtel.ɪ.græf/ [1794] (OED2).

(5) Polysyllabic word with same third syllable: the target is a trisyllabic word sharing the third syllable with the model word, as in *underverse* /ˈʌn.də.vɜːs/ [2008] 'a place gone after death; hell' (RUND) ← *universe* /ˈjuː.nɪ.vɜːs/ [1589] (OED3).

4.2. Morphotactic similarity

On a scale of morphotactic similarity, we can identify an evident similarity in the morphological mechanism, process, or operation obtaining both T and M. This similarity allows for a preliminary taxonomy which may be later refined by identifying internal resemblances between T and M, for instance, concerning compound components, either head or modifier, or blends' splinters. The following classification accommodates analogies into an array of morphological categories and sub-categories ranging from grammatical to extra-grammatical morphological phenomena. An ambiguous case of ungrammatical derivation (*prettiful*) is also discussed in (3).

(1) COMPOUNDING: the target shares with the model the same concatenation operation and either head or modifier. The following patterns describe the analogies met in our database:

(a) N + N compounding with the same head, as in *cakeday* [2013] 'the day on which someone created his/her account' (RUND) ← *birthday* [1574] (OED2);

(b) N + N compounding with the same modifier, as in *buckytube* [1991] 'a cylindrical molecule of carbon' (OED2, [1997] in RUND) ← *buckyball* [1989] 'a molecule of buckminsterfullerene' (OED2);[13]

(c) A + N compounding with the same head, as in *Pink Friday* [2008] 'The Friday after Thanksgiving, on which participating major retailers cut prices and make a donation from sales to help fight breast cancer' (RUND) ← *Black Friday* [1961] (OED3);

(d) name + N compounding with the same head, as in *DianaWorld* [1998] (NEWJT) ← *CharlesWorld* (used in the same micro-context);14

(e) V + Adv compounding with the same second constituent, as in *try-hard* [1922] 'a person who tries very hard' (OED3, [2010] in RUND) ← *die-hard* [1844] (OED2);

(f) A + A compounding with the same head, as in *blank-American* [2008] 'being only American with no other ethnic background' (RUND) ← *African-American* [1835] (OED3), *Mexican-American* [1948] (OED3), etc.;

(g) N + A compounding with the same head, as in *professioncentric* [2008] 'centred in one's profession' (RUND) ← *egocentric* [1900] (OED2);

(h) N-in-N compounding with the same second constituent, as in *buddy-in-law* [2011] 'your friends' friend' (RUND) ← *brother-in-law* [*c*1300] (OED2);

(i) V-and-V compounding with rhyming constituents, as in *wake-and-bake* [2008] 'the process or act of smoking marijuana right upon waking up in the morning' (RUND) ← *Shake-and-Bake* [1965] 'a brand of instant meals' (OED3), with a further similarity in the second verb *bake*.

(2) COMBINING FORM FORMATION: the target shares with the model the same combining form, either a neoclassical combining form of Latin/

Greek origin, such as *-sphere*, or an abbreviated/secreted combining form. Abbreviated ones, such as *e-* (from *electronic*), retain all the semantic components of their full form, whereas secreted ones, such as *-oholic* (from *alcoholic*), retain only some components (e.g. 'a person who is addicted to'), but discard others (e.g. 'alcoholic drinks').[15] The target exhibits:

(a) a neoclassical combining form in *blogosphere* [1999] 'the cultural or intellectual environment in which blogs are written and read' (OED3, [2008] in RUND) ← *-sphere* in *biosphere*, *ecosphere*, etc. and *Diet-sodaphobia* [2008] 'the fear of drinking diet sodas' (RUND) ← *-phobia* in *hydrophobia*, *Anglophobia*, etc. In *technophobe* [1946] 'a person who fears technology' (OED3, [2010] in RUND), an initial combining form (*techno-* in *technology*) combines with a final one (*-phobe* in *hydrophobe*);

(b) an abbreviated combining form in *eco-chic* [1975] 'concern with environmental issues' (OED3, [1997] in NEWJT) ← *eco-*(logical), and in a series of words obtained by *e-*(lectronic) prefixation: e.g. *e-education* [1999] 'education on the web' (NEWJT), *e-reader* [1995] 'a person who reads electronic text' (OED3), *e-shopping* [1998] 'shopping on the web' (NEWJT), *e-text* [1990] 'electronic text' (OED3), *e-voting* [2008] 'online voting' (RUND), and so on. In the case of *e-*formations, well-established words such as *e-mail* [1979] (OED3) and *e-book* [1988] (OED3) constitute the model. The existence of productive series, such as those just illustrated, could also provide a more stable (prefix-like) status to the elements *eco-* and *e-*;

(c) a secreted combining form in words such as *bridezilla* [1995] 'a woman thought to have become intolerably obsessive or overbearing in planning the details of her wedding' (OED3, [2008] in RUND) ← (God)-*zilla*; *cameraholic* [2008] 'one who is addicted to taking pictures' (RUND) and *caffeineaholic* [2010] 'a person who is addicted to caffeine' (RUND) ← (alco)-*holic*; *doctor-speak* [1998] 'the jargon of doctors' (NEWJT) ← Orwell's (New)-*speak*; *coolicious* [2008] 'describing cool with a more modern connotation' (RUND) ← (de)-*licious*; *PDA-athon* [2008] 'a large amount of public display of affection' (RUND) ← (mar)-*athon*; *edutainment* [1983] 'informative entertainment' (OED2, [2008] in RUND) ← (enter)-*tainment*. Some of these combining forms were originally recurrent splinters in blends, but are currently labelled final combining forms in OED3 (see § 2 for the diachronic development of schemas). An initial combining form can be found in *Frankenfood* [1992] 'genetically modified food' (OED3, [1998] in RUND) ← *Franken-*(stein).

In this case, the model words are the earliest formations which exhibit the combining form: e.g., *workaholic* [1947], *chocoholic*

[1961], and *foodaholic* [1965] have established the pattern for *cameraholic*, and *walkathon* [1930] (OED3) and *talkathon* [1934] (OED2) for *PDA-athon*.

(3) Derivation: the target and the model are obtained by adding the same suffix to a comparable base. In this category, however, the target does not refer to an abstract pattern of regular suffixation, but rather to a precise word obtained by rule. Examples of this category include *alphabetism* [1978] 'prejudice or discrimination resulting from a person's position on a (notional) alphabetical list' (OED3, [1997] in NEWJT) ← *racism* [1903], *nanaism* [2008] 'an often humorous error in speech, particularly that made by a grandmother' (RUND) ← *Bushism* [1984] 'a verbal peculiarity or lapse associated with George W. Bush' (OED3); *complify* [2008] 'to render more complex' (RUND) ← *simplify* [1759] (OED2). In *writeo* [2008] 'a misspelling in handwriting' (RUND) ← *typo* [1916] (OED2) the target shares with the model the final vowel *o*, reinterpreted as a suffix (cf. *-o* in *preggo* from *pregnant*; cf. the splinter *-o* in Bauer *et al.* 2013: 527). The humorous neologism *prettiful* [2008] 'having beauty' (RUND) ← *beautiful* [*c*1443] (OED3) is ungrammatical, because an adjectival suffix *-ful* is added to an adjective (*pretty*) rather than to a noun (*beauty*). However, this may also be interpreted as a blend from *pretty* and *beautiful*. A derived word obtained by adding a suffix to an acronymic base is *nimf-ism* [2006] (from *nimf* 'Not In My Front seat' NEWJT) ← *Nimbyism* [1986] (from *Nimby* 'Not In My Back Yard' OED3).

(4) Conversion: the target and the model are obtained by conversion or zero-derivation. For instance, the verb *to Skype* [2003] 'to have a spoken conversation over the Internet using Skype software' (OED3, [2010] in RUND) is obtained from the proprietary name *Skype*, like its model *to Google* [1998] 'to use the Google search engine to find information on the Internet' (OED3) is back-derived from the name of the Internet search engine *Google*.

(5) Blending: like the model, the target is obtained by merging two words into one, specifically:

(a) N + N with overlap (one name is in its full form; see partial blends in Mattiello 2013), as in *Brangelina* [2008/2013] 'Brad (Pitt) + Angelina (Jolie)' (RUND) ← *Bennifer* [2008] 'Ben (Affleck) + Jennifer (Lopez)' (RUND);

(b) N + N with overlap (the first word is shared), as in *sexercise* [1942] 'sexual activity regarded as exercise' (OED3, [1999] in NEWJT) ← *sexpert* [1924] 'an expert in sex' (OED3) (cf. also *sexploitation* [1924] 'sexual exploitation' OED3);

(c) N + N with no overlap (the second word is shared), as in *politisoap* [2005] 'politics + soap' (referred to the television film 'The Deal', NEWJT) ← *docusoap* [1979] 'documentary + soap' (OED3);

(d) N + N with no overlap (the first splinter is shared), as in *brinner* [2008] 'dinner that consists of breakfast food' (RUND) ← *brunch* [1896] 'breakfast + lunch' (OED2);

(e) N + N with/with no overlap (the second splinter is shared), as in *Twittizen* [2010] 'Twitter + citizen' (RUND) ← *netizen* [1984] (OED3). In *bitchdar* [2008] 'bitch + radar' (RUND) ← *gaydar* [1988] 'gay + radar' (OED3), the shared splinter comes from a lexicalised acronym (*radar*).

These sub-categories are not meant to be exhaustive, in that other subcategories of blending exist (see the classification provided by Mattiello 2013: 119-126 along different parameters), but representative of the blending category for analogical formations. Moreover, they differ from the category of clipped compounds because they merge two words that are not commonly treated as a unit, such as *net* and *citizen*.

(6) Clipping: like the model, the target is obtained by clipping part of a word, a compound, or noun phrase, as in *romcom* [1971] 'romantic comedy' (OED3, [2008] in RUND) ← *sitcom* [1964] 'situation comedy' (OED2) and *Indipop* [2008] 'Indian pop' (RUND) ← *Britpop* [1977] 'British pop (music)' (OED3).

(7) Acronymic formation: like the model, the target is obtained by retaining the initial letters of a phrase, a list, or compound, as in *FLOTUS* [2008] 'First Lady of the United States' (RUND) ← *POTUS* [1895] 'President of the United States' (OED3), *OMJ* [2008] 'Oh my Jeez!' (RUND) ← *OMG* [1917] 'Oh my God!' (OED3), or in *ROFL* [2008] 'Rolling On the Floor Laughing' ← *LOL* [1989] 'Laughing Out Loud' (OED3).

4.3. Semantic similarity

On a scale of semantic similarity, we can identify affinities in meaning between T and M or, with complex words, between the free or bound morphemes which T and M consist of. Semantic similarity, therefore, presupposes a comparable morphological make-up of T and M.

(1) Same compound head

(a) with co-hyponym modifier: in endocentric compounds, the target shares the head with the model and the modifiers differ only in a

small set, as in *air-rage* [1996] 'extreme anger or frustration felt during a flight' (OED3, [1998] in NEWJT and RUND) ← *road rage* [1988] (OED3) and *birdhouse* [2011] 'a house or shelter for birds' (NEWJT) ← *doghouse* [1555] (OED3). The second member is shared and the first one is a co-hyponym of the hypernym 'food' in the exocentric compound *beefcake* [1949] ← slang *cheesecake* [1934] (see § 3.2);

(b) with antonym modifier: in endocentric compounds, the target shares the head with the model and the modifiers are opposite, as in *dick-flick*16 [2013] 'movie that favours a male audience' (RUND) ← slang *chick-flick* [1988] 'a film perceived as appealing particularly to women' (OED2) and *long-cut* [2008] 'the long way to get somewhere' (RUND) ← *short-cut* [1619] (OED2);

(c) with quasi-antonym modifier: in endocentric compounds, the target shares the head with the model, whilst the modifier is a quasi-antonym in *vaporware* [1984] 'a piece of software which either does not exist or has not (yet) been developed commercially' (OED3, [2008] in RUND), coined after *hardware* [1947] (OED2, and later *software* [1960], *firmware* [1968], etc.). The same semantic relationship is between the modifier *e-* (from *electronic*, see (2b) in § 4.2) in *e-pal* [2008] 'a friend who communicates with one by e-mails' (RUND) and *pen* in the model *pen pal* (cf. *penfriend*);

(d) with semantically related modifier: in *necromonger* [2008] 'person who spreads death through killing others' (RUND) ← *warmonger* [1590] (OED2), the head is shared and the modifier is less relevantly connected with the model's one.

(2) Same modifier

(a) with co-hyponym head: the modifier is shared with the model, while the head is a co-hyponym of the model's head in *smartwatch* [2013] 'a watch that is worn on a wrist that has connective capabilities' (RUND) ← *smartphone* [1980] (OED3). In the appositional compound *café-bar* [1938] (OED2, [2008] in RUND) ← *café-restaurant* [1926] (OED2), the second component *bar* is a co-hyponym of *restaurant* in the set 'entertainment places';

(b) with opposite head: the modifier is shared between T and M, while the heads are antonyms in *Rice Queen* [2008] 'a non-Asian male who is dating an Asian male' (RUND) ← *Rice King* [2008] 'a non-Asian male who is dating an Asian female' (RUND)17 and *bell-girl* [1998] 'a girl who answers the bell in a hotel' (NEWJT) ← *bell-boy* [1861] (OED2).

(3) SAME PREPOSITION WITH CO-HYPONYM VERBAL BASE: in *walk-off* [2008] 'a competition between models in which the first makes his/her way down a runway in a fashion that is difficult to replicate' (RUND) ← *run-off* [1902] 'an election held to decide the issue between two candidates' (OED3), *walk* and *run* are co-hyponyms of 'verbs of motion'.

(4) SAME SUFFIX WITH CO-HYPONYM BASE: in nominal suffixation, the target base is a co-hyponym (an ethnic adjective) of the model base in *Canadian-ness* [1997] 'the quality or state of being Canadian or of embodying Canadian characteristics' (NEWJT) ← *Englishness* [1804] (OED3). In *Clintonism* [1992] 'the policies or principles advocated by William J.D. Clinton' (OED2, [2008] in RUND) ← *Bushism* [1980] (OED2), the target and model bases are both 'US Presidents' family names'.

(5) SAME BASE WITH ANTONYM PREFIX: in verbal prefixation, the target prefix is antonymous with the model prefix in *prepone* [1913] 'to bring forward to an earlier time or date' (OED3, [1997] in NEWJT) ← *postpone* [1496] (OED3).

(6) SAME SPLINTER AND CO-HYPONYM SPLINTER/WORD: in blending, one of the splinters is shared and the other is a co-hyponym of the model splinter in *Chinglish* [1957] 'Chinese + English' (OED3, [2008] in RUND) ← *Spanglish* [1954] (OED2) and *beefalo* [1974] 'beef + buffalo' (OED2, [2010] in RUND) ← *catalo* [1889] 'cattle + buffalo' (OED2). In *Obamanomics* 'Obama + economics' [2008] (RUND) ← *Nixonomics* [1969] (OED3), *Clintonomics* [1992] (OED2), and in *Romneycare* [2011] 'Romney + healthcare' (RUND) ← *Obamacare* [n.d.] (RUND), a splinter (*-nomics*) or a compound component (*care*) is shared and the other full word is 'the name of a (candidate for) President of the United States'.

(7) QUASI-SYNONYM BLEND COMPONENTS: the two blend components are synonyms denoting 'huge, large in size' in *gimongous* [2010] 'gigantic + humongous' (RUND) ← *ginormous* [1948] 'gigantic + enormous' (OED2). In *superbulous* [2008] 'super(b) + fabulous' (RUND) ← *fantabulous* [1959] 'fantastic + fabulous' (OED2), the blend components indicate something 'excellent, first-class'.

(8) CO-HYPONYM BLEND COMPONENTS: the two blend components are co-hyponyms of 'meals' in *brinner* [2008] 'breakfast + dinner', *linner* [2010] 'lunch + dinner', and *lupper* [2013] 'lunch + supper' (RUND), all from lexicalised *brunch* [1896] (OED2). The components are co-hyponyms of 'a famous couple's name' in both the target *Merkozy* [2011] '(Angela) Merkel + (Nicholas) Sarkozy' (RUND and NEWJT) and the model(s) (*Bennifer* [2008] RUND, *Brangelina* [2008/2013] RUND).

(9) ANTONYM BLEND COMPONENTS: the two blend components are antonyms in *adultescent* [1996] 'adult + adolescent' (OED3, [2003] in NEWJT) ← *kidult* [1960] 'kid + adult' (OED2).

(10) ANTONYM ACRONYM COMPONENTS: the full phrase from which the target acronym originates comprises a component which is opposite to a model's component in *DILF* [2008] 'Dad I'd Like to Fuck' (RUND) ← *MILF* [1992] 'Mother I'd Like to Fuck' (OED3).

4.4. Similarity at various levels

There are some additional cases of analogical neologisms in our database that are more difficult to classify because they can be properly accommodated in more than one scale of similarity. The noun *tri-linguist* /traɪˈlɪŋ.gwɪst/ [1997] 'one who speaks three languages' (NEWJT), for instance, is a quasi-homophone word with its model *bilinguist* /baɪˈlɪŋ.gwɪst/ [1884] (OED2) (see sub-category (2e), phonological similarity in § 4.1), yet the target and model also share morphotactic similarity, because they are obtained by neoclassical prefixation (*bi-* 'two', *tri-* 'three' OED2) of the same base *linguist*.

Another tricky case is the adjective *truthworthy* [1997] 'worthy of the truth' (NEWJT), which is both morphotactically and semantically similar to *trustworthy* [1829] (OED2). From the formal viewpoint, target and model are both adjectival compounds obtained by noun plus adjective concatenation and same syntactic head (see sub-category (1g) in § 4.2). Moreover, from the semantic viewpoint, the two compounds share their head (*worthy*) and the modifiers (*truth* and *trust*) are semantically related abstract nouns (see sub-category (1d) in § 4.3). It should also be observed that *truth* and *trust* are both monosyllabic words with a /tr/ onset, which makes target and model closer from the phonological point of view as well.

For the above-mentioned and similar words, it is more complicated to determine whether the association process relies on one or the other scale, or, more plausibly, if it is the result of a combination of association processes based on resemblances from diverse viewpoints. Indeed, there is often no ambiguity (either one or the other similarity scale), but intended polyvalence (both one and the other scale(s)), especially in sophisticated coinages.

5. Discussion and generalisations on analogical neologisms

The analysis of the neologisms in section 4 demonstrates that it is possible to identify 1) recurrences in the way analogy operates (pro-

ducer's viewpoint) and in the way analogical association works (interpreter's viewpoint), and 2) series of words formed analogically after a schema. Furthermore, the analogical formations analysed in section 4 can be accommodated within the following preliminary morphological categorisation:

– Derivational type: prefixation (*prepone* ← *postpone*); suffixation (*Canadian-ness* ← *Englishness*, *prettiful* ← *beautiful*);

– Compound type: with same head (*blamestorming* ← *brainstorming*, *blank-American* ← *African-American*, *professioncentric* ← *egocentric*); with same modifier (*bellgirl* ← *bellboy*);

– Particle compound type: *walk-off* ← *run-off*;

– Reduplicative compound type: *wake-and-bake* ← *Shake-and-Bake*;

– Conversion type: *to Skype* (from *Skype*) ← *to Google* (from *Google*);

– Combining form type: neoclassical combining form (*technophobe* ← *techno-* in *technology* + *-phobe* in *hydrophobe*), abbreviated combining form (*eco-chic* ← *ecological*), secreted combining form (*bridezilla* ← *Godzilla*, *coolicious* ← *delicious*);

– Blending type: total blend (*linner* 'lunch + dinner' ← *brunch*); partial blend (*Brangelina* 'Bra(d) + Angelina' ← *Bennifer*); with overlap (*sexercise* 'sex + exercise' ← *sexpert*, *adultescent* 'adult + adolescent' ← *kidult*); with no overlap (*brinner* 'breakfast + dinner' ← *brunch*);

– Clipping type: *romcom* 'romantic comedy' ← *sitcom*;

– Acronym/initialism type: *FLOTUS* ← *POTUS*, *OMJ* ← *OMG*.

This categorisation confirms that surface analogy recurs throughout the spectrum from rule-based to extra-grammatical formations. It operates both in regular grammatical morphology (derivation, compounding, conversion) and in extra-grammatical morphology (blending, clipping, acronym formation) (Mattiello 2013). On the other hand, analogy via schema especially works in the sub-module of marginal morphology – namely, in transitional phenomena between derivation and compounding (Dressler 2000: 7), represented by the combining form type.

Analogy, therefore, is a promising area of investigation, which should not be neglected on account of its relevance to lexical enrichment and vocabulary expansion. Many new words are nowadays obtained by analogy with existing ones – e.g. more than 11% in NEWJT, with 21.05% recent neologisms, and 5.89% in RUND, with 12.56% recent neologisms – and many of these neologisms also conform to rules. Analogy, therefore, is neither unrelated to productivity nor antonymic with rules. In NEWJT, surface analogies conforming to rule patterns are nearly one sixth of the total number of analogies and 62.10% are analogies via schema formed from an established combining form. In RUND, regular surface analogies are 16.58%, with a 27.27 percentage of recent neologisms. Among analogies via schemas, 73.84% exhibit well-established combining forms and 12.82% are attested headwords in the OED.

This quantitative data corroborates the assumption that novel analogical formations can both be based on precise lexical items, and obey regular patterns of concatenation and productive rules. In other words, the fact that *alphabetism* is associated with the model *racism* and *nanaism* is specifically created after *Bushism* does not exclude the fact that they exhibit a productive suffix *-ism*.

Overall, in the two collections, surface analogy is more frequently illustrated by blends and acronyms, surface analogy also conforming to rules by compound words, and analogy via schema by combining form combinations. These tendencies have been confirmed by the findings in a larger project on the subject of analogy in new English words (Mattiello, *in preparation*).

The present analysis has a twofold function: first, a synchronic study of the database shows the morphological categories and subcategories that are relevant to analogical formation, disconfirming that analogy is only related to creative extra-grammatical formations and demonstrating its pertinence to productive grammatical words as well. Second, a diachronic study of those model words that have triggered a series can help predict new targets obtained after the same schema.

Moreover, the associations identified in section 4 for the various scales show that the similarity relationship between targets and models involve phonological and semantic factors, besides morphosyntactic ones. Some of these associations also appear to be more frequent than others, allowing for generalisations which are, nevertheless, not as stable as those based on productive rules. In particular:

– Phonological association especially occurs when the target and the model are quasi-homophones differing only for a vowel (e.g. *Internot* ← *Internet*) or a consonant sound (e.g. *ego-tourism* ← *eco-tourism*). However, from the morphotactic viewpoint, *ecotourism* combines an abbreviated form *eco-* (from *ecological*) with a full base, whereas *ego-tourism* is a regular nominal compound;

– Morphotactic association instead occurs when, for example, in compounds such as *professioncentric*, after *egocentric*, the target and the model have their head in common, although the two modifiers are not semantically related. It may also occur when two compounds share their modifier (e.g. *buckytube* ← *buckyball*), or when two blends share the first or second splinter (e.g. *brinner* ← *brunch*, *Twittizen* ← *netizen*). In analogy via schema, the association is with a series of words, as in *blogosphere* ← *-sphere* in *biosphere*, *ecosphere*, etc. and *e-education* ← *e-* in *e-mail*, *e-book*, etc.;

– Semantic association finally occurs when, for example, in compounds, derived words, or extra-grammatical formations, the components of the target and the model are related by a relationship of similarity (*gimongous* ← *ginormous*), opposition (*prepone* ← *postpone*, *long-cut* ← *short-cut*), or co-hyponymy (*air-rage* ← *road rage*, *linner/lupper* ← *brunch*);

– Lastly, more than one classificatory feature may increase the similarity between target and model, and make the association process more straightforward. In *truthworthy*, the association with *trustworthy* is driven by similarity at different levels of analysis. Simultaneous associations of the target with the model at different levels can facilitate the disambiguation of the former, and the recoverability of the latter.

The similarity scales proposed show different degrees of similarity on various levels (phonological, morphotactic, and semantic) which may intermingle. In these scales, preferred models are those which resemble their targets from various viewpoints (e.g. *trustworthy*), so that the model-target association is immediate and, consequently, the model recoverability is favoured. Preferred models are also series (e.g. *e-mail*, *e-book*, etc.), in that the existence of accepted sets of words sharing the same formation process is another factor helping and encouraging model-target association. Finally, models that are complex (or re-analysable as complex words) may be preferred in analogical formation (e.g. *Inter-* + *net*, *ego* + *centric*, *post-* + *pone*, although the verb **pone* does not exist in English, cf. Latin), in that this may help

the substitution (i.e. *net* with *not*, *ego* with *profession*, *post-* with *pre-*) that is typical of the analogical proportion.

Other generalisations that can be made concern blends and combining forms. Some blends exhibit repeated components. For instance, *sex* frequently occurs as a first component in overlapping blends whose second component begins with the onset /eks/ (*sexercise* [1999] NEWJT, *sexcellent* [2008], *sexcursion* [2008], *sexile* [2008], *sexperiment* [2008], *sextortion* [2013] RUND, all analogical with *sexpert*, *sexploitation*). The splinter *docu-*, as in *docusoap*, based on *docudrama*, is another case in point. Therefore, it is likely that future blends obeying the same patterns are coined, and that *sex-* or *docu-* become initial combining forms, shifting from type (2a) (more unstable) analogy to type (2b) (stable) analogy via schema (see § 2). This is indeed the way *-tainment* [1990], or *-licious* [1878], or *-zilla* [1978] have become established combining forms (OED3).

It is even plausible to envisage shifts from surface analogies, based on precise words, to analogies via schemas, in that single cases like *blamestorming*, analogous to *brainstorming*, may establish a schema suiting future coinages, such as **faultstorming* or **headstorming*, which are also analogous with either target or model from the semantic viewpoint. Thus, what is currently target (e.g. *blamestorming*) may become the model for future coinages.

The coinage of analogous formations is often, especially at their early creation, an allusion to the model word: for instance, the target *Clintonism* alludes to the adherence to or support of the policies or principles advocated by Clinton (as in the model *Bushism*). This case may suggest similar targets, such as *Obamaism*, which is nowadays attested on *Google* and in COCA, although not in dictionaries. Here and in the previous case of *blamestorming*, the targets entirely comply with derivation and compounding rules, and this facilitates the creation of analogical neologisms based on the same pattern.

Journalistic language as found in NEWJT and partially in RUND is particularly rich in neologisms created by analogy with jocular existing ones. In The Guardian online,[18] a journalist has created some new blends on the pattern of *Merkozy*: i.e. *Cleggeron* '(Nick) Clegg + (David) Cameron', *Balooper* '(Ed) Balls + (Yvette) Cooper', *Blush* '(Tony) Blair + (George) Bush', and others. Although not all of these formations will become true neologisms attested in dictionaries, some of them may, and this would confirm the power of analogy in word-formation.

6. Conclusions

Analogy is a process that marks items of partial similarity by association. The three scales of similarity identified and illustrated in this paper can account for the association of target words with models at different language levels. Ranging a new target word along one of these scales means attributing to it a classificatory feature belonging to the model, and hence categorising the former on the basis of the latter.

This paper has shown that the establishing of scales with different degrees of resemblance between targets and models can contribute to a better understanding of the analogical mechanism. From the viewpoint of the model, a high degree of resemblance with the target can facilitate its recoverability, i.e. the identification of the word or word set which acted as model. From the viewpoint of the target, a high degree of resemblance with the model can encourage the association and ease the process of disambiguation, that is, the understanding of the new target word.

These claims can be supported by results obtained from psycholinguistic experiments on native English speakers, including both online processing tests and offline tests. So far, only offline tests have been conducted on the predictability and acceptability of new English analogical words (Mattiello, *in preparation*), demonstrating that similarity between model and target can encourage target acceptability and facilitate model recoverability, both when the target is provided in isolation and, especially, when it is in co-textual occurrence with the model. Other factors motivating the acceptability of analogical formations are their regularity and conformity also to rule patterns, and their triggering series. However, online priming tests with lexical decision tasks and related reaction times could also be used in future research to corroborate these findings.

Furthermore, the scales of similarity that have been kept distinct in the analysis may often intersect in the association process. Thus, *gimongous* 'gigantic + humongous' can be associated with its model *ginormous* 'gigantic + enormous' because the blend components are near synonyms (semantic similarity), but also because the first blend splinter *gi-* is shared (morphotactic similarity), and because the two blends also share the first unstressed syllable and the rhyme in the last syllable. In target words which show manifest likeness to their models at various levels, such as *gimongous* or *truthworthy*, there is an intended polyvalence (both one and the other scale(s)) which activates the association.

From the morphological categorisation of the analogies in our database, we have verified that analogies recur throughout the spectrum from rule-based to extra-grammatical formations. The model of analogy that we have discussed in this paper is gradual and ranges from surface analogy, based on specific lexical items (*DILF* ← *MILF*, *prepone* ← *postpone*), to analogy via schema, based on prototype words which constitute a series (*e-text* ← *e-mail*, *e-fit*, etc.). Depending on the number of attested words which follow the same pattern, we have also distinguished between more stable *vs* less stable schemas, the former being represented by combining forms (e.g. -(*o*)*holic* in *sexholic* or *eco-* in *eco-chic*) and the latter by, for instance, repeated blend splinters, such as *-nomics* (from *economics*) in *Nixonomics*, *Clintonomics*, and *Obamanomics*.

Although for analogy generalisations are not as stable as generative-like ones, based on productive rules, we can establish different degrees of productivity in analogical formations, from more productive patterns in analogy via schema of the -(*o*)*holic* type to less or unproductive patterns in surface analogy obtained by extra-grammatical operations (i.e. the *DILF* type).

Finally, we can regard analogy via schema as an extension of surface analogy. Indeed, especially when the latter conforms to productive rule patterns, it may become the model for new analogies. We can, for instance, envisage a series of formations from ethnic adjectives created after *Canadian-ness* (e.g. *Italian-ness*, *French-ness*, *German-ness*, etc.), or nouns ending in *-ism* with American Presidents' names as bases. Indeed, it is debatable whether *Clintonism* is a surface analogy created on *Bushism*, or rather comes from a productive pattern also including *Nixonism* [1952] (OED3) and *Reaganism* [1966] (OED3). *Obamaism* could, in any case, be added to this list of *isms*.

This study is limited to a restricted number of neologisms drawn from existing collections. The analysis of a larger database is needed to corroborate our findings and give more precise results (Mattiello, *in preparation*). Given the relevance of neologisms to journalistic and young people's language as well as to specialised terminology and literary works, it is our intention to expand our research on analogy towards these directions.

Notes

[1] Throughout the paper, the years provided in square brackets refer to the earliest attestations of the words either in the OED (second or third edition – OED2/OED3), or in other specified sources.

[2] The term 'assimilation' has been avoided in the present paper – and 'association' has been preferred – because it is typically used in phonology. Thus, its use in a different context may be misleading.

[3] For the origin and history of the notion of analogy, see Hill (2007), Schironi (2007), and Rainer (2013). For a more succinct account, see Dressler & Laaha (2012).

[4] See also Hock's (1991: 171) "four-part analogy" and its diachronic relevance to the morphological regularisation of some English plurals (e.g. *cow-s vs* ME *kine*) or past tenses (e.g. *help/help-ed/help-ed vs* ME *help/holp/holpen*).

[5] In inflectional morphology, schemas are understood in a very restricted sense, i.e. they are only relevant to isolated paradigms and families of paradigms (Dressler & Laaha 2012: 49).

[6] In the OED, the first quote for model and target is the same: *viz.*, "[1539] *One Eye wytnesse, is of more value, than tenne eare witnesses*".

[7] According to OED2, *illfare* was coined in opposition to *welfare* to refer to 'the condition of faring or getting on badly; infelicity; adversity'. The earliest attestation of the word in OED2 is in 1425, but it has later [1962] been revived and used in phrases such as *the Illfare State* (cf. *Welfare State*) for jocular effects (see also its use in the book title *Illfare in India*).

[8] For poetic audacity and neologisms in literary texts, see Dressler (1981, 1993), Neuhaus (1989), Iamartino (1999), Ladányi (2000), Boase-Beier (2010), and Merlini Barbaresi (2011).

[9] The APRIL project (http://rdues.bcu.ac.uk/april.shtml) is concerned with the development of a system for the semi-automatic classification of rare words in journalistic text, over a period of years. See also the RDUES web page http://rdues.bcu.ac.uk/neologisms.shtml.

[10] Some of the entries come from a previous collection, *Neologisms* (2003), but the current RUND has started to exist in this form since 2008, with approximately 5,500 words at the beginning and more than 3,500 new entries added until 2013. The most recently added terms are dated December 2013.

[11] Both archives go back to 2004 and a subscription is necessary to widen the research to previous years, which are however irrelevant to current neologisms.

[12] See Kilani-Schoch & Dressler (2005) for a more homogeneous parameterisation in French sub-regular verbs.

[13] The chemical word *bucky* is a back-clipping from *buckminsterfullerene* (OED2).

[14] "Yet, just a year later, there can no longer be any doubt that, if one chooses to see Britain simply as a battlefield for a war between CharlesWorld and DianaWorld, CharlesWorld is winning decisively". (NEWJT 1998).

[15] For a classification of English and Italian combining forms, see Mattiello (2008).

[16] According to OED2, the noun *Dick* is a familiar pet-form of the common Christian name *Richard*, and hence it is generically used to refer to a 'fellow, lad, man'.

[17] Although target and model are added in the same year to the RUND database, the compilers specify that *Rice Queen* is coined after *Rice King*.

[18] See the article *Forget Brangelina, it's time for Merkozy* appeared on 24th October 2011 at the address www.theguardian.com.

Bibliographical References

Anttila, Raimo 1977. *Analogy*. The Hague, Paris & New York: Mouton.
Aronoff, Mark 1976. *Word Formation in Generative Grammar*. Cambridge: MIT Press.
Aronoff, Mark 1983. Potential words, actual words, productivity and frequency. In Hattori, Shiro & Inoue, Kazuko (eds.), *Proceedings of the 13th International Congress of Linguistics*. Tokyo: Permanent International Committee on Linguistics. 163-171.
Aronoff, Mark & Fudeman, Kristen A. 2011 [2004]. *What is Morphology?* Malden (MA), Oxford & Chichester: Wiley-Blackwell.
Baldi, Philip & Dawar, Chantal 2000. Creative processes. In Booij, Geert E.; Lehmann, Christian; Mugdan, Joachim; Kesselheim, Wolfgang & Skopeteas, Stavros (eds.), *Morphologie-Morphology: An International Handbook of Inflection and Word-formation*. *Vol. I*. Berlin & New York: Walter de Gruyter. 963-972.
Bauer, Laurie 1983. *English Word-formation*. Cambridge: Cambridge University Press.
Bauer, Laurie 1988. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
Bauer, Laurie 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
Bauer, Laurie; Lieber, Rochelle & Plag, Ingo 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
Becker, Thomas 1990. *Analogie und morphologische Theorie*. München: Fink.
Blevins, James P. & Blevins, Juliette 2009. Introduction: Analogy in grammar. In Blevins, James P. & Blevins, Juliette (eds.), *Analogy in Grammar. Form and Acquisition*. Oxford: Oxford University Press. 1-12.
Bloomfield, Leonard 1933. *Language*. New York: Henry Holt & Company.
BNC = *British National Corpus* 2010. BNC Webmaster. University of Oxford. http://www.natcorp.ox.ac.uk/.
Boase-Beier, Jean 2010 [1987]. *Poetic Compounds: The Principles of Poetic Language in Modern English Poetry*. Tübingen: Max Niemeyer Verlag GmbH & Co KG.
Booij, Geert E. 2005. *The Grammar of Words*. Oxford: Oxford University Press.
Booij, Geert E. 2010. *Construction Morphology*. Oxford: Oxford University Press.
Brinton, Laurel J. & Traugott, Elizabeth C. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
Christofidou, Anastasia 1994. *Okkasionalismen in poetischen Texten*. Tübingen: Narr.
Clark, Eve V. 2009 [2003]. *First Language Acquisition*. Cambridge: Cambridge University Press.
COCA = *Corpus of Contemporary American English* 1990-2012. Brigham Young University. http://corpus.byu.edu/coca/.

De Smet, Hendrik 2013. Change through recombination: Blending and analogy. *Language Sciences* 40. 80-94.
Dressler, Wolfgang U. 1981. General principles of poetic license in word formation. In Geckeler, Horst; Schlieben-Lange, Brigitte; Trabant, Jürgen & Weydt, Harald (eds.), *Logos semantikos. Studia linguistica in honorem Eugenio Coseriu 1921-1981. Vol. II*. Berlin: Mouton de Gruyter. 423-431.
Dressler, Wolfgang U. 1993. Word-formation: Poetic licence. In Asher, Ronald E. (ed.), *The Encyclopedia of Language and Linguistics*. *Vol. IX*. Oxford: Elsevier Pergamon. 5028-5029.
Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. In Doleschal, Ursula & Thornton, Anna M. (eds.), *Extragrammatical and Marginal Morphology*. München: Lincom Europa. 1-10.
Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics / Rivista di Linguistica* 15 (1). 31-62.
Dressler, Wolfgang U. & Laaha, Sabine 2012. The impact of types of analogy on first language acquisition. *Wiener Linguistische Gazette* 76. 48-62.
Dressler, Wolfgang U. & Ladányi, Mária 1998. On grammatical productivity of word formation rules (WFRs). *Wiener Linguistische Gazette* 62/63. 29-55.
Fertig, David L. 2013. *Analogy and Morphological Change*. Edinburgh: Edinburgh University Press.
Fischer, Olga 2007. *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: Oxford University Press.
Fradin, Bernard 2000. Combining forms, blends and related phenomena. In Doleschal, Ursula & Thornton, Anna M. (eds.), *Extragrammatical and Marginal Morphology*. München: Lincom Europa. 11-59.
Gardani, Francesco 2013. *Dynamics of Morphological Productivity. The Evolution of Noun Classes from Latin to Italian*. Leiden: Brill.
Google. 2016. www.google.it.
(The) Guardian online. 2004-2016. Guardian News and Media Ltd. www.theguardian.com/uk.
Hermann, Eduard 1931. *Lautgesetz und Analogie*. Berlin: Weidmannsche Buchhandlung.
Hill, Eugen 2007. Proportionale Analogie, paradigmatischer Ausgleich und Formerweiterung. Ein Beitrag zur Typologie des morphologischen Wandels. *Diachronica* 24 (1). 81-118.
Hock, Hans H. 1991. *Principles of Historical Linguistics*. Berlin & New York: Mouton de Gruyter.
Hockett, Charles F. 1968. *The State of the Art*. The Hague & Paris: Mouton.
Hopper, Paul J. & Traugott, Elizabeth C. 2003 [1993]. *Grammaticalization*. Cambridge: Cambridge University Press.
Iamartino, Giovanni 1999. L'innovazione lessicale nei testi letterari e nelle loro traduzioni: Puntualizzazioni concettuali e indicazioni metodologiche. In Azzaro, Gabriele & Ulrych, Margherita (eds.), *Anglistica e…: Metodi e percorsi comparatistici nelle lingue, culture e letterature di origine europea*. *Vol. II*. Trieste: EUT. 257-269.
(The) Independent online. 2004-2016. Independent Print Ltd. www.independent.co.uk.

Kilani-Schoch, Marianne & Dressler, Wolfgang U. 2002. Affinités phonologiques dans l'organisation de la morphologie statique: L'exemple de la flexion verbale française? *Folia Linguistica* 36 (3-4). 297-312.

Kilani-Schoch, Marianne & Dressler, Wolfgang U. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Narr.

Kiparsky, Paul 1974. Remarks on analogical change. In Anderson, John M. & Jones, Charles (eds.), *Historical Linguistics*. Amsterdam: North Holland. 257-275.

Klégr, Aleš & Čermák, Jan 2010. Neologisms of the 'on-the-pattern-of' type: Analogy as a word-formation process? In Procházka, Martin; Malá, Markéta & Šaldová, Pavlína (eds.), *The Prague School and Theories of Structure*. Göttingen: V&R unipress. 229-241.

Köpcke, Klaus-Michael 1993. *Schemata bei der Pluralbildung im Deutschen: Versuch einer kognitiven Morphologie*. Tübingen: Narr.

Krott, Andrea 2009. The role of analogy for compound words. In Blevins, James P. & Blevins, Juliette (eds.), *Analogy in Grammar. Form and Acquisition*. Oxford: Oxford University Press. 118-136.

Ladányi, Mária 2000. Productivity, creativity and analogy in word formation (WF). Derivational innovations in Hungarian poetic language. In Alberti, Gábor & Kenesei, Istvaìn (eds.), *Approaches to Hungarian: Papers from the Pécs Conference*. *Vol. VII*. Szeged: JATEPress. 73-90.

Lahiri, Aditi 2000. Introduction. In Lahiri, Aditi (ed.), *Analogy, Levelling, Markedness. Principles of Change in Phonology and Morphology*. Berlin & New York: Mouton de Gruyter. 1-14.

Lehrer, Adrienne 1996. Identifying and interpreting blends: An experimental approach. *Cognitive Linguistics* 7 (4). 359-390.

Lehrer, Adrienne 2003. Understanding trendy neologisms. *Italian Journal of Linguistics / Rivista di Linguistica* 15 (2). 369-382.

Lehrer, Adrienne 2007. Blendalicious. In Munat, Judith (ed.), *Lexical Creativity, Texts and Contexts*. Amsterdam & Philadelphia: Benjamins. 115-133.

Lieber, Rochelle 2005. English word-formation processes. Observations, issues, and thoughts on future research. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of English Word-formation*. Dordrecht: Springer. 375-427.

Marchand, Hans 1969. *The Categories and Types of Present-day English Word-formation*. München: Beck.

Marle, Jaap van 1990. Rule-creating creativity: Analogy as a synchronic morphological process. In Dressler, Wolfgang U.; Luschützky, Hans C.; Pfeiffer, Oskar E. & Rennison, John R. (eds.), *Contemporary Morphology*. Berlin & New York: Mouton de Gruyter. 267-273.

Mattiello, Elisa 2008. From *sexgate* to *vallettopoli*: Contrasting English and Italian combining forms. In Vergaro, Carla (ed.), *Dynamics of Language Contact in the Twenty-first Century*. *Vol. II*. Perugia: Guerra. 177-190.

Mattiello, Elisa 2013. *Extra-grammatical Morphology in English. Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin: Mouton de Gruyter.

Mattiello, Elisa *in preparation*. Analogy in Word-formation. A Study on

English Neologisms and Occasionalisms. Submitted manuscript.
McMahon, April M. S. 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
Merlini Barbaresi, Lavinia 2011. Outstanding features of the Italian language: Evaluatives, alteratives, blends, and compounds in the Italian rendition of Joyce's *ALP*. In Baccolini, Raffaella; Chiaro, Delia; Rundle, Chris & Whitsitt, Sam (eds.), *A Joyceful of Talkatalka from Friendshapes for Rosa Maria Bollettieri Bosinelli*. *Vol. I*. Bologna: Bononia University Press. 195-216.
Miller, Gary D. 2014. *English Lexicogenesis*. Oxford: Oxford University Press.
Motsch, Wolfgang 1981. Der kreative Aspekt in der Wortbildung. In Lipka, Leonhard (ed.), *Wortbildung*. Darmstadt: Wissenschaftliche Buchgesellschaft. 94-118.
Neologisms. 2003. Rice University. www.ruf.rice.edu/~kemmer/Words04/neologisms/a.html.
NEWJT = *Neologisms – New Words in Journalistic Text*. 1997-2012. Birmingham City University. http://rdues.bcu.ac.uk/neologisms.shtml.
Neuhaus, H. Joachim 1989. The Shakespeare dictionary database. *ICAME Journal* 13. 3-11.
Oxford English Dictionary Online. 1989-currently updated. Second Edition (OED2) / Third Edition (OED3). Oxford: Oxford University Press. www.oed.com.
Paul, Hermann 1880. *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer.
Plag, Ingo 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin & New York: Mouton de Gruyter.
Plag, Ingo 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
Rainer, Franz 2013. Formación de palabras y analogía: Aspectos diacrónicos. *Anexos de Revista de Lexicografia* 19. 141-172.
RUND = *The Rice University Neologisms Database*. 2004-2014. Rice University. http://neologisms.rice.edu/.
Ronneberger-Sibold, Elke 2000. Creative competence at work: The creation of partial motivation in German trade names. In Doleschal, Ursula & Thornton, Anna M. (eds.), *Extragrammatical and Marginal Morphology*. München: Lincom Europa. 85-105.
Ronneberger-Sibold, Elke 2008. Word creation. Definition - function - typology. In Rainer, Franz; Dressler, Wolfgang U.; Kastovsky, Dieter & Luschützky, Hans C. (eds.), *Variation and Change in Morphology. Selected Papers from the 13th International Morphology Meeting*, Vienna, 3-6 February, 2008. Amsterdam & Philadelphia: Benjamins. 201-216.
Scalise, Sergio 1988. The notion of 'head' in morphology. In Booij, Geert E. & Marle, Jaap van (eds.), *Yearbook of Morphology*. Dordrecht: Kluwer. 229-245.
Schironi, Francesca 2007. *Analogía*, analogia, proportio, ratio: Loanwords, calques, and reinterpretations of a Greek technical word. In Basset, Louis; Biville, Frédérique; Colombat, Bernard; Swiggers, Pierre &

Wouters, Alfons (eds.), *Bilinguisme et terminologie grammaticale gréco-latine*. Louvain: Peeters. 321-338.
Skousen, Royal 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer Academic Publishers.
Skousen, Royal 2009. Expanding analogical modeling into a general theory of language prediction. In Blevins, James P. & Blevins, Juliette (eds.), *Analogy in Grammar. Form and Acquisition*. Oxford: Oxford University Press. 164-184.
Szymanek, Bogdan 2005. The latest trends in English word-formation. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-formation*. Dordrecht: Springer. 429-448.
Traugott, Elizabeth C. & Heine, Bernd, eds., 1991. *Approaches to Grammaticalization. 2 Vols*. Amsterdam & Philadelphia: Benjamins.
Wanner, Dieter 2006. *The Power of Analogy. An Essay on Historical Linguistics*. Berlin & New York: Mouton de Gruyter.
Wikipedia. 2001-2016. Wikimedia Foundation, Inc. https://www.wikipedia.org/.
Zemskaja, Elena A. 1992. *Slovoobrazovanije kak dejatel'nost'* [Word-formation as Activity]. Moskva: Nauka.