Studia Anglica Posnaniensia 53 (2018): 67–114
doi: 10.2478/stap-2018-0004

THE MORPHOSEMANTIC TRANSPARENCY/OPACITY OF NOVEL ENGLISH ANALOGICAL COMPOUNDS AND COMPOUND FAMILIES[1]

ELISA MATTIELLO,[2] WOLFGANG U. DRESSLER[3]

University of Pisa, University of Vienna

ABSTRACT

This study deals with novel English analogical compounds, i.e. compounds obtained via either a unique model (e.g. *beefcake* after *cheesecake*) or a schema model: e.g., *green-collar* based on *white-collar*, *blue-collar*, *pink-collar*, and other X-*collar* compounds. The study aims, first, to inspect whether novel analogical compounds maintain the same degree of morphosemantic transparency/opacity as their models, and, second, to find out the role played by the compound constituents in the constitution of compound families, such as X-*collar* and others. To these aims, the study proposes a scale of morphosemantic transparency/opacity for the analysis of compound constituents. In particular, the compound constituents in our database (115 examples) are analysed in connection with: 1) their degree of transparency (vs. opacity, including metaphorical/metonymic meaning), linked to their semantic contribution in the construction of the whole compound's meaning, and 2) their part-of-speech. Against the common assumption that productive word-formation rules mostly create morphosemantically transparent new words, or that rule productivity is closely connected with transparency, the study of our database demonstrates that novel analogical compounds tend to maintain the same transparency/opacity degree as their models. It also shows that, in nuclear families and subfamilies of compounds, the part-of-speech of the constituents, their degree of transparency/opacity, and their semantic relation are reproduced in all members of the analogical set.

Keywords: Morphosemantic transparency/opacity; Analogy; Analogical compounds; Compound family

---

1 The data used for this paper was collected during a research project on analogy in English neologisms carried out partly at the University of Vienna and partly at the University of Pisa. The project, covering a two-year period (2014-2016), was financed by a NetWordS (European Network on Word Structure) grant, which allowed for a collaboration between the authors.

2 Department of Philology, Literature and Linguistics, University of Pisa.

3 Austrian Centre for Digital Humanities and Austrian Academy of Sciences, University of Vienna.

1. Introduction

Analogical compounds are compounds obtained by analogy. The latter is a word-formation process whereby new words are created that are coined either on precise actual model words ("surface analogy", Mattiello 2016, 2017a, after Motsch's 1981: 101 "Oberflächenanalogie"), or after a schema model ("analogy via schema", Mattiello 2017a; cf. Bybee 1988), i.e. after a series or word family.

Our view of schema differs from the conception developed by Booij (2010). Within Booij's Construction Morphology, schemas come in two types. A non-productive schema captures patterned relations among listed instances, but resists extension to new instances. A productive schema also captures patterned relations among listed instances, but in addition can be used freely to create new instances. It is the latter function that corresponds most closely to traditional productive rules. As to the relationship between analogy and abstract schemas in word-formation, within Construction Morphology (Booij 2010), Booij claims that they are opposite endpoints on a scale of schematicity. In other words, he argues that there is no absolute boundary between analogy and abstract schemas, denying that the latter may be viewed as a concrete notion. As for composition, he identifies patterns of compounding by constructional subschemas of various degrees of abstraction and generalisation, arguing their necessity on the basis of observations on semantic specialisation, headedness variation, diachrony, and allomorphy selection (Booij 2010).

In this paper, by contrast, a schema is defined as a concrete model identifiable as two or more target words that provide a pattern for a set of formations (see Köpcke 1993). The words that serve as schema consist of either a series, i.e. a homogeneous set of prototype actual words sharing the same formation, or a family of words sharing (some of) the bases.

The notion of word family is actually more complex and has evolved over time in different directions. Originally, a word family was defined as consisting of "a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (Bauer & Nation 1993: 253). Later (e.g. in Bertram, Baayen & Schreuder 2000; De Jong *et al.* 2002; Hay & Baayen 2002) the concept was extended to word sets sharing the base, or one of the stems in compounds. Hence, the concept of 'compound family' was established to denote a word family made up of similar compounds that share at least one of their components in the same constituent position.

In this study on analogical compounds and compound families, we agree with Booij (2010: 93) that "[t]he formation of new compounds is not necessarily based on the model of existing compounds". In order to be categorised as analogical with (i.e. based on) an existing compound or compound family, a new compound

has to meet such criteria as 1) sharing one constituent with the model (Invariable Part); 2) sharing the communal constituent position; and 3) following the model(s) chronologically.[4] The role of 4) the transparency/opacity degree of the compound constituents, 5) their syntactic pattern, and 6) their semantic relation will be the object of analysis in this study.

Our hypothesis is that, besides the Invariable Part, also the non-shared constituent (Variable Part) in analogical compounds should reproduce the model, both semantically (same transparency/opacity degree) and syntactically (same part-of-speech). This similarity would indeed help the association of a novel compound to its model. However, there are cases of analogical compounds which seem to contradict this hypothesis. For instance, the noun *blamestorming* [1997] 'the process of investigating the reasons for a failure and of apportioning blame' (OED3), which according to the OED is obtained after *brainstorming* [1907] 'the action or process of making a concerted attempt to solve a problem' (OED3), does not replicate its model semantically. Indeed, the Variable Part of the target *blame* is more transparent than metaphorical *brain* in the model and, as a consequence, the semantic relation between the target and model compound constituents also differs. However, in this case, the phonological resemblance between *blame* and *brain* can be an indicator of analogy (see Mattiello 2016 for types and scales of similarity in analogical neologisms). In general, we believe that this issue should be addressed more systematically and in quantitative terms, in order to determine the role played by the transparency/opacity of compound constituents in the coinage of novel analogical compounds and (nuclear) compound families.

In particular, this study investigates both compounds obtained by surface analogy, which have a unique model, and compound families, which have triggered a schema model for existing or potential analogical compounds. For instance, compounds such as *white-collar* [1911] 'relating to non-manual work' (OED3), *blue-collar* [1929] 'relating to manual work' (OED3), *pink-collar* [1975] 'relating to employment associated with women' (OED3, s.v. *pink*), and *green-collar* [1992] 'designating work relating to the preservation of the environment' (OED3, s.v. *green*) have established the pattern for the compound family having *collar* as second stem (e.g. *black-collar*, *gold-collar*, *gray-collar*, etc. in Benczes 2006: 144-145). Thus, X-*collar* provides an instance of nuclear compound family whose second constituent (*collar*) is shared and whose first constituent belongs to a restricted set of adjectives designating a type of colour (e.g. *white*, *blue*, *pink*, *green*, etc.). The compounds also share the same A–N

[4] For each compound, the earliest attestation in the OED will be provided in square brackets. The compound's meaning will be offered according to the second (OED2) or third updated edition (OED3).

pattern and a comparable semantic relation between their constituents. Therefore, a nuclear compound family is different from an extended compound family, including, for the above-mentioned case, *dog collar* [1485] or *horse-collar* [1497], with a different N–N pattern and a completely unrelated meaning. By contrast, *collar-bone* [1615] does not belong to the extended compound family either, in that the constituent *collar* does not share the same (head) position (cf. "reverse families" in Bell & Schäfer 2016: 168), nor does it share the syntactic A–N pattern, or the meaning 'relating to a type of work' with the above nuclear family. As Booij (2010) remarks, the existence of constituent families is validated by the "Family Size Effect", establishing that the larger is the size of a constituent family, the faster it will be retrieved in a lexical decision task (De Jong *et al.* 2002). Whereas psycholinguistic research is far from the interests of this study, we will investigate the role of morphosemantic transparency/opacity degree, part-of-speech, and semantic relation between constituents in the identification of nuclear family members within semantically homogeneous sets of compounds.

The study aims, first, to inspect whether novel analogical compounds maintain the same degree of morphosemantic transparency/opacity as their models, and, second, to find out the role played by compound constituents in the constitution of compound families. To these aims, the study proposes a scale of morphosemantic transparency/opacity for the analysis of compound constituents (cf. Schwaiger *et al.* 2017 for German diminutives; a much smaller one by Talamo, Celata & Bertinetto 2016 for Italian derivatives). In particular, the compound constituents are analysed in connection with: 1) their degree of transparency (vs. opacity, including metaphorical/metonymic meaning), 2) their part-of-speech, and 3) their semantic contribution in the construction of the whole compound's meaning (cf. Frege's "Principle of Compositionality", in Partee *et al.* 1990 *inter alia*).

In this study, the notion of 'morphosemantic transparency' is applied to constituent transparency and viewed in terms of both meaning relatedness, i.e. how the meaning of a compound constituent word is related to the main meaning of the same word used individually, and meaning predictability, i.e. how the meaning of a compound constituent word can contribute to the overall meaning of the compound (to its compositionality). In these terms, a compound is fully compositional when both constituents are morphosemantically transparent and their relation can be regularly predicted.

In the analysis, the following research questions will be addressed:

1) Do novel analogical compounds maintain the same transparency/opacity degree as their models? If not,
   a) Is there a tendency towards more transparency in the target than in the model?

   b) How can the model be retrieved?
2) How can we distinguish a nuclear compound family from an extended family? More specifically,
   a) What is the role of the constituents' part-of-speech in the distinction?
   b) What is the role of the semantic relation between the compound constituents?

As a more general goal, the study expects to identify morphosemantic patterns according to which novel analogical compounds can be created or understood. In addition, it expects to find that nuclear compound families, whose members are morphosemantically very close to one another, play a much more important role than extended families in the formation and interpretation of novel analogical compounds. For the selection of compounds, advanced search in the OED and in relevant online collections of neologisms was integrated with manual search in online sources.

Another general goal is to disprove in detail the general assumptions that: (1) productive word-formation rules mostly create morphosemantically transparent new words (e.g. Aronoff 1976: 22, 32-33, 38-39), (2) rule productivity is intimately connected with transparency (e.g. Bauer 2001: 54, 60; Plag 2003: 177; Gardani 2013: 92), or (3) only or mainly diachronic change results in opacity (e.g. Aronoff 1976: 18-19; Bauer 1983: 48-50, 55-59). Some doubts have been cast on these assumptions (e.g. Bauer, Lieber & Plag 2013: 14; Rainer *et al.* 2014: 8), but these have not been thoroughly investigated with compounds so far. Note that all the neologistic compounds that we are discussing here have been formed by productive compounding rules (in Section 2 we explicitly exclude the very rare case of surface analogies which are not based on productive rules).

The hypothesis upon which our study rests is that, while for novel compounds, which are not analogically formed but motivated by other cotextual or contextual factors, we may envisage a tendency towards semantic transparency to help the interpreter in disambiguation (Schäfer 2018: 13, 24), for novel analogical compounds the tendency may be to keep either transparency or opacity of the constituents. It is our aim to demonstrate that analogy may justify the coiner's choice of less transparent or even opaque constituents for new compounds, provided that the semantic similarity with the model is maintained. When it is not – i.e. when the degree of transparency is increased or decreased – the motivation for the change may be phonological, or connected to ease of processing (Gagné & Spalding 2014; Libben & Weber 2014). This is another supposition that motivates a fine-grained investigation of a set of novel analogical compounds.

2. The modern notion of analogy and its relevance to word-formation

The modern notion of analogy goes back to Neogrammarians (from Paul 1880 onwards), who adopted the Greek proportional descriptive technique (A : B = A' : X, X = B') to describe, for example, plural formation in English, as in *cup* : *cups* = *pot* : X, obtaining *pots*. Bloomfield (1933) considered this type of analogy as a paradigmatic substitution in a proportional equation, e.g., *pot* is substituted for *cup* in the above-mentioned proportion.

Against the American generativist tradition (Chomsky 1957; Aronoff 1976), where analogy was exclusively viewed as a local mechanism not allowing for generalisations, the main defenders of proportional analogy (e.g. Hockett 1968) also applied the concept to derivation, bringing it back to its ancient eminence. Then analogy was adopted in psycholinguistics (Goswami 1991) and in computational (exemplar-based) models (Skousen & Stanford 2007). Recently, it has been considered one of the leading principles in language learning and language change (Anttila 2003; Fertig 2013).[5]

Currently, Mattiello (2016) has focused the attention on the role played by analogy in word-formation, showing that this process is transversely relevant to grammatical, "extra-grammatical" (Mattiello 2013), and "marginal" morphology (Dressler 2000). Within extra-grammatical morphology, analogy applies to the creation of words whose input and formation mechanisms are not clearly identifiable by word-formation grammar, and which do not allow a prediction of a regular output. For instance, the blend *blaxploitation* [1972] 'the exploitation of black people' (OED2) is clearly modelled on the precise word and word form of *sexploitation* [1924], by merging *black* with *exploitation*. This is called "pure surface analogy", i.e. analogy whose crucial motivation is pure similarity with a concrete model and not with an abstract pattern (Mattiello 2017a). By contrast, when "surface analogy combines with rule patterns" (Mattiello 2017a), the motivation is twofold, namely, similarity with a precise lexical item and conformity to derivation or compounding rules.[6] The latter type is in the formation of regularly derived words, such as *alphabetism* [1978] 'prejudice or discrimination resulting from a person's position on an alphabetical list', after *racism* [1903] and *sexism* [1906], but also obeying to *-ism* suffixation. Here the meaning conveyed by the models – i.e. 'belief in the superiority of one race/sex over another' – is more specific than the sense commonly conveyed by the suffix *-ism*. This same sense of 'prejudice or

---

5 For an overview on the concept of analogy, see the recent account by Arndt-Lappe (2015).

6 When models are formed by a non-productive rule, new analogical words are rare, with the exception of occasionalisms created by audacious writers, poets, or in public advertisements (cf. Dressler & Tumfart 2017).

discrimination' is also in analogical *ageism* [1969] 'discrimination against the elderly' (OED3), *sizeism* [1971] 'discrimination against those viewed as overweight or short' (OED3), and *speciesism* [1975] 'discrimination against certain animal species by human beings' (OED2). This sense is so common nowadays that it has become an additional meaning of the suffix itself.

Surface analogy combines with rule patterns also when it creates regular compounds, as in the oft-quoted example of *ear-witness* [1539] 'a person who testifies to something that he or she has heard' (OED3), formed after *eyewitness*, earliest attested in 1539 in the OED, but clearly preceding its analogical formation.

Hence, analogy is a diachronically relevant concept. First, it is historically that we can determine the 'model' word and the resulting analogy (called 'target'), the latter being more recent than its model. Second, the diachronic evolution of language can govern changes, for instance, from a blend's "splinter" (Lehrer 1996, 2003; Bauer, Lieber & Plag 2013) to a combining form, or even to a suffix, given its productivity in terms of frequency, regularity, and predictability in the formation of new words. As Plag (2003: 38) admits, in cases such as *-burger* – from the reanalysis of *hamburger* as *ham* + *burger*, as in *cheeseburger* [1938] – analogy can give rise to productive word-formation rules. Indeed, *-burger* has given birth to a productive series (cf. Bauer 1983: 96), including, besides *cheeseburger*, also *chickenburger* [1936], *beefburger* [1940], *vegeburger* [1945], *eggburger* [1960], and similar words (more examples in Bauer, Lieber & Plag 2013: 526). The process behind these formations is, therefore, "analogy via schema", i.e. based on a set of *-burger* words.

Bauer, Lieber & Plag (2013: 519) claim that this type of processes are part of "paradigmatic morphology" (as pioneered by van Marle 1985), i.e. based on some sort of similarity between words in the lexicon. In particular, they claim that:

> In compounding, there are many forms that are modelled on particular existing compounds, with the new compound inheriting crucial components of the institutionalized meaning of the model compound. (Bauer, Lieber & Plag 2013: 524)

Novel English analogical compounds which are modelled on existing compounds are the focus of attention in this study. We believe that the 'crucial components' that are inherited by analogical compounds have to do with the morphosemantic transparency/opacity of the constituents, their position within the compound, their word class, and their reciprocal semantic relation.

3. Status of research of analogical compounds, compound families, and their semantic transparency

Compounds are very common in English, but often confused with lexical phrases (also called multi-lexical words), which we have excluded from the present study. According to Lieber (2005: 376-377), the criteria that are used to distinguish the former from the latter are spelling, lexicalised meaning, stress, and inseparability of the first and second stems. Actually, some of these criteria are disconfirmed by many actual instances. For example, most compounds can accept different spellings (*daisy wheel*, *daisy-wheel*, *daisywheel*), they can be stressed on the left or right stem (*blackboard* vs. *apple pie*), or can be lexicalised or compositional in meaning (*blackboard* vs. *whiteboard*). However, all noun compounds are immune to separation by a modifier (e.g. **black wooden board*), which is often true also for multi-lexical words and idiomatic phrases. Thus, we believe that no absolute reliable criterion, but a combination of all the above probabilistic criteria can be used for determining compounds in English.

In this study, we adopt a wider view of compounds, also including "loose compounds" (Scalise 1992), which, unlike "strict compounds", do not allow for phonological amalgamation, but exhibit inseparability.[7] A relevant example in English is analogical *white márket* [1943] 'authorized dealing in commodities that are rationed' (OED3, s.v. *white*), which, like its model *black márket*, can be labelled compound for its inseparability, in spite of its atypical right hand stress. On the other hand, our study excludes compounds that are not analogy-based: e.g., *sunflower* is not modelled on another compound, but is formed by combining two freestanding words. In other words, there is no English compound denoting a type of flower which shares the second constituent with and may have acted as model for *sunflower* [1562] (cf. *night-flower* 1648). In this case, *sun-flower* is a literal translation of Greek *hēli-anthous*, which has inspired the creation of an English compound with no indigenous model.

Compounds, whose constituents are freestanding words, also have to be kept distinct form combining forms (e.g. *-logy* in *biology* or *-holic* in *workaholic*), splinters (e.g. *-ercise* in *sexercise*), or affixes (e.g. *-ism* in *racism*), which typically cannot stand in isolation and are likewise excluded from our interests in this study.

The existing literature on analogical compounds mainly focuses on the psycholinguistic relevance that these compounds have to first language acquisition, stress assignment, processing, and interpretation of novel

---

7 With regard to inseparability, cf. Bell (2012), in which the author shows that inseparability does not hold for all compounds that are standardly considered to be compounds.

compound words (cf. Krott 2009 for the production, interpretation, and acquisition of noun-noun compounds). A pioneer study in the field of the acquisition of English morphology is Berko (1958). Her results show that small children have a tendency to coin new words by using compounding patterns that they already know. For instance, they create **zibman* for 'a man who **zibs* for a living', after the compounds *postman* [1529] or *milkman* [1589] (although from a verbal base), and *baby* **wug* for 'a very tiny **wug*', after the regular pattern of *baby elephant* [1815] or *baby bird* [1841].

Another relevant and more recent block of psycholinguistic studies has conducted experiments on the effects of constituent families on stress assignment in novel English compounds. In particular, Plag (2010) has demonstrated that stress assignment in N^1–N^2 compounds is largely predictable from the stress behaviour of related compounds that have the same left or right constituent. Thus, compounds with the head *street* tend to be left-stressed (*Máin Street*, *Óxford Street*), whereas compounds with the head *avenue* or *lane* are rather right-stressed (*Fifth Ávenue*, *Madison Ávenue*, *Oxford Láne*). Interestingly, Bell & Plag (2013) have also shown that constituent families may be associated with different stress patterns depending on the semantics of the construction involved: cf. *toy fáctory* 'a model factory for playing with' vs. *tóy factory* 'a factory that makes toys'. This clearly shows the importance of semantic relations in compound families.

In the same field of research, Arndt-Lappe & Bell (under revision) have recently adopted Skousen & Stanford's (2007) analogical algorithm "AM::Parallel" to model stress assignment in a corpus of 486 nominal compounds. In line with the authors, stress is assigned to new compounds in accordance with the stress pattern of similar compounds previously encountered and stored in the lexicon. In their version of the AM model, degree of similarity is calculated using the compounds' constituents and semantic properties. Results from their study actually show that a large number of compounds with similar semantics have the same stress behaviour, but, for many other compounds, stress assignment is based on more local analogues: e.g., *lamb sándwich* and *salmon sándwich* provide a small analogical set for stress assignment in the new compound *banana sándwich*. Hence, our distinction between surface analogy vs. via schema.

Moreover, in psycholinguistic studies on the processing and interpretation of novel compound words, compound constituent families have been found (see Libben 2008) to be the locus of forming new compounds added by analogy. As for compound processing, Libben (2008) has argued that it is characterised by "Maximization of Opportunity". In other words, when processing a compound, whole-word activation combines both with constituent activation and with morphological proliferation. Thus, for example, for the morphological parsing of

a new compound such as *black-collar* (see § 5), the whole word is activated, its constituent free morphemes (*black* and *collar*) are activated, and an additional set of positionally bound morphemes (*black-* and *-collar*) are also activated. Compound family activation – i.e. the activation of the families including *blackmail*, *blackboard*, etc. and *white-collar*, *blue-collar*, etc. – can contribute, in this approach, to the interpretation of the new compound. Experiments on "Family Size effect" (De Jong *et al.* 2002; Gagné 2009: 262-263) have confirmed the view that novel compound words having a high Family Size (in terms of type frequency of their constituents), either in the non-head or in the head component, are easier to process than those having a low Family Size. This study will demonstrate that the activation of nuclear compound families and subfamilies is more relevant than the activation of extended families for novel analogical compounds. For instance, *blackboard* cannot be the model for *black-collar* because the field it belongs to (i.e. school) is totally unrelated to the meaning of the latter novel compound. Meaning, therefore, and especially the same degree of transparency/opacity of the compounds' constituents and their contribution to the sense of the whole compound, can be viewed as discriminatory criteria distinguishing a nuclear family from an extended one.

Still other experiments test the interpretation of novel N–N compounds based on prototypical semantic relations between components. For instance, Smith, Barratt & Zlatev (2014) have studied compound food names such as *Parma ham* positing that, since the default interpretation for such compounds involves physical origin, this interpretation is also activated with novel compounds, with rare exceptions (e.g. *Hawaii pizza*, with pieces of ham and pineapple, has a Canadian origin).

Compound processing has finally been studied in connection with semantic transparency and morphological headedness (Gagné 2009: 264-268). Psycholinguistic studies have underlined the centrality of semantic transparency in the processing of English compounds. For instance, Libben (1998, 2010) has investigated the role of morphological decomposition in the processing of semantically transparent vs. opaque compounds. In Libben (1998) and related studies, the author classifies N–N and A–N compounds in terms of degree of constituent transparency: i.e., transparent–transparent (TT), as in *doorbell*; partially transparent ("partially compositional" in Bourque 2014: 115), i.e., opaque–transparent (OT, e.g. *strawberry*) or transparent–opaque (TO, e.g. *jailbird*); and fully opaque (OO), as in *humbug*. Results show that constituent activation occurs for both transparent and opaque compounds. Notably, Jarema *et al.* (1999: 362) have used this study as a point of departure to demonstrate that "the semantic transparency of individual constituents, their position in the string, and morphological headedness interact in the processing of compounds" (see also Gagné & Spalding 2014; Libben & Weber 2014).

In compound words, semantic transparency is a concept that is generally viewed as a matter of constituents' meaning. That is, "[a] compound word is usually defined as transparent when the meaning of the compound word is consistent with the meanings of the constituents (e.g., *carwash*). In contrast, a compound word is defined as semantically opaque, when its meaning cannot be constructed by directly combining the meanings of the individual constituents (e.g., *pineapple*)" (Pollatsek & Hyönä 2005: 262).

However, as remarked in psycholinguistic studies and recently highlighted by Bourque (2014: 2), the binary opposition between transparent or opaque is not sufficient to describe compounds. For instance, many compounds involving the same lexemes show considerable differences at the level of meaning construal. Thus, while a *housefly* 'a fly typically found in houses' is fully transparent and *gadfly* 'an annoying person' is fully opaque, compounds such as *firefly* 'a nocturnal beetle that emits light', *butterfly* 'an insect with large, colourful wings', and *barfly* 'a person who spends much time in a bar' are neither fully transparent nor fully opaque in their overall meaning. However, in *barfly*, *bar* is transparent and, in *firefly*, *fire* is figuratively used, since it metonymically stands for 'light'. Hence, we need a more granular approach to the concept of "semantic transparency", which should not be conflated with "compositionality", although the former clearly depends on the latter (see Bourque 2014: 40-46 for a clear distinction between the two concepts). In this study, we will propose our scale of morphosemantic transparency (differing from Bourque's typology) according to which compounds' constituents can be rated.

More precisely, the typology elaborated by Bourque (2014) is based on four basic factors, namely: 1) headedness (endo- vs. exo-centricity), 2) compositionality (i.e. how individual constituents contribute meaning to the whole), 3) implicit semantic relations within compounds, and 4) semantic homogeneity (i.e. the degree of shared meaning between analogically similar compounds).

Bourque's (2014: 295-297) typology incorporates all of these features into a hierarchy consisting of sixteen possible configurations based on headedness and compositionality. Of these possible transparency profiles, he found that only twelve were relevant in French.

As for semantic relations, in his typology of semantic transparency of French compounds, Bourque (2014: 276-291) has proposed a distinction between fully compositional (i.e. fully transparent), weakly compositional, partially compositional, and non-compositional, i.e. totally opaque compounds (cf. degrees 1 and 6 in § 5). He defines compositionality as "determined according to individual components' meaning in relation to that of the whole" (Bourque 2014: 258). Thus, in his typology, strongly endocentric compounds can be fully, weakly or partially compositional, whereas only exocentric compounds can be

non-compositional. Bell & Schäfer (2016), instead, is a recent study presenting models of semantic transparency in which "the perceived transparency of English noun–noun compounds, and of their constituent words, is predicted on the basis of the expectedness of their semantic structure" (p. 157). Moreover, Schäfer (2018) has just published his work on the semantic transparency of English compound nouns showing the importance of compound family based models. In this work, the author concludes that "all semantic-based predictors reflect […] expectancies drawn from the distribution of the respective features across the compounds' positional constituent families" (Schäfer 2018: 266). In particular, for the assessment of compounds' semantic transparency, Schäfer (2018) explores two factors: i.e. the semantic relations holding between compound constituents and the role of different readings of the constituents and the whole compound, operationalised in terms of the distribution of specific readings across constituent families. However, Schäfer (2018) considers semantic transparency as a binary rather than a scalar concept, and his set of nominal compounds, mainly drawn from COCA and other online resources, does not necessarily consist of novel compounds.

Although "[e]xtensive descriptive work has been undertaken on the semantic relations holding between the components of English compounds" (Fabb 1998: 74; see, e.g., Levi 1978), and some of them are thought to be cognitively more accessible than others, these relations are not pertinent for a scale of semantic transparency.

In our study, we adopt different criteria to assess the semantic transparency of novel English compound words. In particular, our scale of morphosemantic transparency (§ 5) refers to the analysis of the individual compound constituents rather than to the compound as a whole. Thus, our classification (§§ 6.1-6.7) partially intersects (but does not overlap) with Bourque's (2014) criteria. For instance, our fully transparent compounds (§ 6.1) are endocentric and fully compositional, in that both constituents exhibit optimal transparency (degree 1–1). However, metonymic or metaphorical compounds (§§ 6.5-6.6) can display figurative meaning in only one constituent (e.g. the modifier), and be both endocentric and partially compositional, according to Bourque's (2014) parameters. As Bourque (2014) admits when discussing metaphor and metonymy in compounds, "[t]he number of potential combinations of tropes in a given compound makes it extremely difficult not only to offer an exhaustive set of features that might affect semantic transparency, but also to determine which of these combinations has the greatest impact" (p. 119).

Furthermore, our goal is different from Bourque's (2014) one. Indeed, this study does not aim to classify English compounds in line with their semantic transparency, but it rather shows how the degree of semantic transparency (vs. opacity vs. figurative meaning) of individual constituents, their part-of-speech,

and semantic relation are of fundamental importance when forming or interpreting a novel analogical compound. Hence, we claim that 1) the semantic relation held between the constituents is generally reproduced in analogical compounds, 2) the semantic homogeneity of a set of complex words can lead to the formation of nuclear compound families, and 3) the similarity of a novel compound to a nuclear compound family can help interpret it according to the same established pattern.

In previous research (Mattiello 2016, 2017a), it has been demonstrated that analogical compounds, such as *blue-collar*, after *white-collar*, or *small fish* [1836] 'a person of small importance' (OED3, s.v. *small*), after *big fish* [1827] 'an important or influential person' (OED3, s.v. *big*), can be described in terms of Variable Part (e.g. *blue* vs. *white*, *small* vs. *big*) and Invariable Part (e.g. *collar*, *fish*). The Variable Part in target compounds is generally semantically related to the corresponding part in model compounds. Semantic similarity includes relations of:

a) Near identity, or even true synonymy: e.g., *mouse race* [2003] 'lower-stress life-style that results from moving to a smaller community or taking a less demanding job' (*Wordspy*), modelled on colloquial *rat race* [1937] 'urban working life regarded as an unremitting struggle for wealth, status, etc.' (OED3) (with *mouse* and *rat* being similar animals distinguished for their size) (for true synonymy, see *big gun* and *great gun* in § 5);
b) Contradictory opposition (polarity antonyms): e.g., *hot war* [1947] 'an armed conflict' (OED3, s.v. *hot*), modelled on *cold war* [1945] 'hostilities short of armed conflict' (OED2, s.v. *cold*) (with an opposition between the gradable adjectives *hot* and *cold*);
c) Contrary opposition (simple or converse antonyms): e.g., *airwoman* [1910] 'a woman who is engaged in the flying or operation of aircraft' (OED3), modelled on its male counterpart *airman* [1873] (with the binary pair *man–woman*), or *househusband* [1858] 'a husband or male partner who carries out the household role and duties traditionally associated with a housewife' (OED3), on *housewife* [*c*1225] (with converse *husband–wife*);
d) Other contrasts (esp. co-hyponymy): e.g., *white-collar*, *blue-collar*, *pink-collar*, *green-collar*, etc. (§ 1) (with co-hyponymy among *white*, *blue*, *pink*, and *green*).

By contrast, the Invariable Part is always morphosemantically identical, with the exception of word plays or puns, i.e. words that are purposefully formed by altering existing words, generally with the intention of playfulness. For

example, the analogical word *herstory* [1970] 'in feminist use: history emphasizing the role of women or told from a woman's point of view' (OED2) is described by the OED as a punning alteration of *history* [OE], reinterpreted as a compound *his–story*. Thus, in *herstory*, the Invariable Part *story* stands for 'history', as if *history* were a haplological form of *his history* (a novel explanation that we propose for such cases).

We expect that a morphosemantic analysis of model and target compound constituents can help investigate analogical compounds in terms of formation, interpretation, and availability for still novel formations.

4. Dataset and methodology

The dataset used for the morphosemantic analysis includes 115 English compounds extracted from online and paper dictionaries, and collections of English neologisms. Given the drawbacks of the *Oxford English Dictionary* as a source for lexical and linguistic research evidenced by recent studies (Mugglestone 2005; Brewer 2014), our data was collected from heterogeneous sources which are considered reliable and thorough for studying new English lexicon. They include:

- Online dictionaries, such as the online version of the *Oxford English Dictionary*, continuously updated from the second (OED2) to the third edition (OED3), or the *Urban Dictionary* (UD), constantly expanded with new items;
- Paper dictionaries, such as John Algeo's (1991) *Fifty Years among the New Words: A Dictionary of Neologisms, 1941-1991*;
- Online sources and web sites, e.g., *Wordspy – The Word Lover's Guide to New Words*, a site created to keep track of emerging vocabulary of the English language;
- Existing collections of new words, namely *Neologisms – New Words in Journalistic Text* (1997-2012) (819 entries) and the *Rice University Neologisms Database* (2004-2014) (9,016 entries).

Online resources such as UD or *Wordspy* are continuously integrated with new examples. This guarantees that the dataset used for the analysis is inclusive enough to provide a realistic picture of current English novel words. Moreover, new entries in UD are rated by using an online system, thus allowing a distinction between idiosyncratic words and widely accepted or recognised new words. For this study, we selected compounds which had been positively rated by native speakers or entered by more than one dictionary maker. A sample of fifteen analogical compounds was also submitted to several native speakers of

British and American English. They were asked their opinion on the semantic transparency of the compound constituents, provided that they knew the novel compounds. In particular, they were asked to rate the transparency of the compound constituents compared with their main meaning when used as independent words. With the exception of three or four compounds that they did not know, they agreed with our ratings on more than 90% of the compound constituents.

The methodology used for data collection included both advanced search and manual selection. Advanced search options and search tools were available for nearly all online resources. For instance, the *Rice University Neologisms Database* offers an advanced search option which allows us to specify the word's grammatical category, word-formation type, and where the submitted word should be found ('anywhere', 'in definition', or 'in source'). The *Urban Dictionary* allows for a search by entry, but also provides cross-references to formally or semantically related words, which may be either the model or the target of analogical formation.

New vocabulary added to the *Wordspy* website was also monitored during our project on analogy. Like UD, *Wordspy* offers both a manual search option, by providing an alphabetically ordered list of the words, and, under each entry, a 'Some Related Words' link, which allows users to compare the currently explored entry with other related ones. Information about the words' etymology is additionally provided under the entries. The OED also offers an advanced search option, whereby words created 'after the word X' can be selected. An exclusively manual search was instead necessary for the paper dictionary (Algeo 1991) and the collection *Neologisms – New Words in Journalistic Text*. The focus of our search was on analogical compounds sharing one of the constituents (Invariable Part).

5. Scale of morphosemantic transparency (vs. opacity)

The scale of morphosemantic transparency elaborated for the analysis of compound constituents is partially adapted from Schwaiger *et al.* (2017). The authors have elaborated a scale of morphosemantic transparency/opacity for the analysis of German diminutives which is finer-grained and more realistic than the binary opposition between transparency and opacity used in the literature for compounds' representation and processing (e.g. Libben 1998). Transparency and opacity are indeed gradable concepts ranging from maximum transparency to total opacity, with intermediate degrees which also involve figurative meaning. For this study, we have envisaged a six-step scale of morphosemantic transparency/opacity (with a further subdivision into 4a and 4b in degree 4), as reported in Table 1. In the analysis, only one constituent of a compound is

considered individually and not the constituents in their interrelation (for the psycholinguistic importance of each constituent's semantic transparency in compound processing, see Gagné & Spalding 2014). Moreover, as competition between potential meanings affects the interpretation of compound constituents, as well as their semantic transparency, we have taken into consideration only the main meaning of the constituents, because an account of all the minor meanings of a constituent as an autonomous word would have resulted in an explosion of varieties to classify and thus endangered any generalisations (Schmidtke *et al.* 2016). Moreover, the main meaning of a word has a good chance to come first into the mind of listeners or readers and thus be the basis of their comparison with the meaning contribution of the respective constituent to the global meaning of the compound. Clearly our scale is based on the semantics of compound-internal relations and not of compound properties (cf. Bell & Schäfer 2016).

Table 1. Scale of morphosemantic transparency/opacity

Degrees	Properties	Examples of English analogical compound
1	Same meaning/Optimal transparency	café-bar 1–X
2	Slight semantic restriction	kiteboard 2–X
3	Strong semantic restriction	slow food 3–X
4	a. Metonymic relation	black-collar X–4a
	b. Metaphorical relation	blamestorming X–4b
5	Weak semantic relation	eternity leave 5–X
6	No semantic relation/Full opacity	big gun X–6

The scale in Table 1 represents a continuum rather than separate degrees, so the proposed degrees of transparency/opacity represent best points with fuzzy boundaries.

Optimal transparency (degree 1) is maximum transparency, as in the analogical compound *café-bar* [1938], obtained by analogy with the appositional compound *café-restaurant* [1926] (both OED2, s.v. *café*). In both model and target, the left constituent *café*, from French, plainly refers to 'a coffee-house' (main meaning), although it has recently acquired also the meaning of 'a restaurant where simple and usually quite cheap meals are served'.

In transparency degree 2, we have a slight semantic restriction, as in the left constituent *kite* of the compound *kiteboard* [1998] 'a type of surfboard designed for riding across water while harnessed to a large kite controlled by hand-held

strings' (OED3), after *surfboard* [1798] (OED3). This is actually a case of immediate abbreviation, from *kite*(*surf*)*board*, where *kite* has a slightly more specific use than the traditional toy consisting of a light frame with paper or other light thin material stretched upon it.

In transparency degree 3, we have a stronger semantic restriction, as in the left constituent *slow* of *slow food* [1974] 'food prepared in a conventional or traditional manner' (OED3), which opposes to its model *fast food* [1954] 'the type of food served in a fast-food restaurant or which can be prepared quickly at home' (OED2). The meaning associated with *slow* in this compound is not exactly that of not quick, ready, or prompt, but has to do with conventionality in culinary traditions. Hence, it is much more specialised than the main meaning of the adjective used individually. Meaning restriction or specialisation, as we will see (e.g. in § 6.3), also includes slang meaning or technical language, and is often related to word polysemy.

Transparency degree 4 is related to figurative use of language. In particular, we have identified two possible interpretations of compound constituents, namely, metonymic meaning and metaphorical meaning. In the scale, metonymic meaning has been attributed a higher degree of transparency (4a) than metaphorical meaning (4b) because in the Idealised Cognitive models elaborated within Cognitive Linguistics by Lakoff & Johnson (1980), the conceptual mappings occurring in metaphor and metonymy involve different degrees of complexity. In particular, metaphor is described as a conceptual mapping (a set of correspondences) from a source domain (vehicle) to a target domain (tenor) (e.g. *LOVE IS A JOURNEY* in *Look how far we've come*) (Lakoff & Johnson 1980). On the other hand, metonymy is reference to an entity in a schema by referring to another entity in the same schema (e.g. *ORDER FOR CUSTOMER* in *The ham sandwich is waiting for his check*) (Lakoff & Johnson 1980). In other words, while metonymy is a domain-internal conceptual mapping between a domain and one of its subdomains (or vice versa), in metaphor the cognitive process involves two different domains, and therefore a higher degree of semantic complexity. Thus, in *black-collar* [2012] 'an unknown independent working artist who toils long hours for relatively little or speculative pay' (*Urban Dictionary*), metonymy links the right constituent *collar* to 'the clothing used by the artist when working', specifically, a part of the clothing stands for the whole clothing, which in its turn stands for the worker (*PART FOR WHOLE* metonymy). By contrast, in *blamestorming*, obtained after *brainstorming* (§ 1), the shared right constituent activates the metaphor *HEATED DEBATE IS STORMING*. One might object that metonymy is syntactically more complex than metaphor. However, for our scale, semantic (not syntactic) dimensions are relevant.

A high degree of opacity is in degree 5, with a weak semantic relation of the compound constituent. For example, the constituent *eternity* in *eternity leave* [1999] (*Wordspy*), after *maternity leave* [1919] (OED3, s.v. *maternity*), is weakly linked with the compound meaning 'paid leave given to a person who needs to provide full-time care for a dying family member'. In this case, *eternity* may be connected with the idea of endlessness of life after death, or it may otherwise refer to a very long period spent with one's family member who is close to death. More plausibly, *eternity* has been chosen as a substitute for *maternity* not for its semantics, but rather for its phonological similarity with the model: i.e. three syllables and the prosodic structure are shared by the two Variable Parts.

Lastly, full opacity is in degree 6, with total lack of semantic motivation. A relevant case is the constituent *gun* in *big gun* [2001] 'an important or powerful person' (OED3, s.v. *big*), after its synonym *great gun* [1657] (OED3). Closeness (identity) between the model and target's Variable Parts (*great* and *big*) helps the association here, and the interpretation of the novel compound.

6. Qualitative (morphosemantic) analysis

In this section, we apply the scale of morphosemantic transparency described and illustrated in the previous section for the analysis of analogical compounds vis-à-vis their model compounds. The primary goal of the analysis is to investigate whether target analogical compounds maintain the same degree of transparency/opacity as their models in their constituents. The constituents under exam for morphosemantic analysis may be either the right or the left components, and either the Variable or the Invariable Parts.

6.1. Fully transparent (endocentric) compounds

Fully transparent analogical compounds are compounds whose constituents both belong to degree 1 of the morphosemantic scale (1–1). Fully transparent also implies fully compositional, in that all constituents contribute semantically to the meaning of the whole. Besides the above-mentioned *ear-witness*, after *eyewitness* (§ 2), other N–N compounds belong to this type. Some cases with a shared head are *moonquake* [1906] 'a seismic tremor of the moon's surface' (OED3), after *earthquake* [*c*1325], *sandboard* [1992] 'a long narrow board on which a rider may coast down sand dunes' (OED3), after *snowboard* [1983], and *father-substitute* [1938] 'a person who assumes the role of a father' (OED3, s.v. *father*), after *mother-substitute* [1933]. The analogical compound *puppy leave* [2000] 'time taken off work to care for a new puppy' (*Wordspy*) similarly maintains the same degree of transparency as its model *maternity leave* [1919]

(cf. opacity in *eternity leave*, § 5). In *Father's Day* [1908] 'a day on which fathers are particularly honoured' (OED3, s.v. *father*), after *Mother's Day* [1874], the constituents are linked by a Saxon genitive. In *small data* [2012] 'the data generated by an individual or small company' (*Wordspy*), after *big data* 'data of a very large size' (OED3, s.v. *big*), the morphosemantic analysis is 1–1, in spite of the restricted use of the compound to sector-based language.

Fully transparent analogical compounds with a shared non-head component are, besides *café-bar*, after *café-restaurant* (§ 5), the noun *podcatching* [2004] 'checking for and downloading any new programs that appear on a podcasting feed' (*Wordspy*), after *podcasting* [2004], and the verb *prooflisten* [2000] 'to listen to a recording of words or music to check for errors' (*Wordspy*), after *proofread* [1845].

The rhyming compounds *handie-talkie* [1942] 'a lightweight walkie-talkie radio set, easily carried in one hand' (OED3) (shortened from *hand*(*walk*)*ie-talkie*) and *walkie-lookie* [1946] 'a portable television camera which transmits pictures and sound wirelessly' (OED3), both obtained after *walkie-talkie* [1939], though respectively sharing the right and the left component, have transparent components, but are exocentric, in that their head 'device' is not expressed (Mattiello 2013: 155). This type of compound shows that there is not always direct correlation between constituent transparency and compound endocentricity. Indeed, this is comparable to the type *redskin*, with transparent constituents but no head 'man' expressed.

In addition, the three-member compound *third-hand smoke* [1991] 'particles that linger on surfaces after second-hand tobacco smoke has dissipated' (*Wordspy*), is analysable as [[1–3]–1], like the model *second-hand smoke* [1891]. Similarly, in *second-hand drinking* [1994] 'a negative effect that a drinker has on a non-drinker' (*Wordspy*), on *second-hand smoking* [1891], the analysis is still that of a compound containing a subordinate compound [[1–3]–1].

An adjectival N–A compound that deserves attention is *dairy-free* [1983] 'that does not contain milk or products derived from milk' (OED2, s.v. *dairy*), based on *sugar-free* [1924], *gluten-free* [1927], etc. (cf. antecedent *tax-free* [1705], which cannot be the model because it belongs to a completely different semantic field, see § 3). The X-*free* compound family is very extensive nowadays, as *-free* has been available for compounding from the early 19th century onwards (Liu & Zhan 2015), but has recently developed a specific meaning related to a health benefit in eliminating some foods, such as sugar, gluten or dairy products (cf. independent *free* or the *free of* and *free from* phrasal constructions). Thus, in *dairy-free*, the first constituent is fully transparent, while the second one is slightly more specific in meaning (degree 2).

6.2. Converted compounds from phrases

There is a group of analogical compounds that are obtained from phrases, via conversion, with word-class (syntactic) more than semantic change. In most of these cases, the component's transparency is not endangered.

For instance, from phrasal verbs, we have *callback* [1914] 'an instance of returning a person's telephone call' (OED3) and *fax-back* [1988] 'a service that can fax a document automatically on request' (OED3), both after *ringback* [1895]. Although the covert head of these compounds is outside (exocentric), the overt constituents are morphosemantically transparent (1–1).

A comparable analysis (degree 1–1) is in *walk-in* [1943] 'a cinema, shop, bank, etc. entered on foot' (OED3), after *drive-in* [1937], and in *sleep-in* [1965] 'a form of protest in which the participants sleep overnight in premises which they have occupied' (OED2), after *sit-in* [1937]. Although these appear to be cases of surface analogy, because the verbs' meaning is strictly related – i.e. *walk* and *drive* are motion verbs, *sleep* and *sit* are verbs of state – the latter example is part of a larger family where *-in* is considered a suffix or converted particle 'indicating any group protest or large gathering for some common purpose' (OED2). Therefore, in *sleep-in* and *sit-in*, *-in* is hard to connect to the meaning of the preposition *in* (degree 5). Examples given in the OED are very numerous: *read-in* [1961] 'a protest at which demonstrators gather to read', *fish-in* [1964] 'a form of protest by American Indians against the loss of fishing rights', *be-in* [1967] 'a public gathering of hippies', *love-in* [1967] 'a gathering at which people are encouraged to express feelings of friendship and physical attraction', etc. The lexical expansion of this group seems to be diachronically related to the 1960s, all after the model *sit-in*.

Another noun compound from a phrasal verb is *warm-down* [1951] 'a period of moderate physical activity undertaken to aid recovery from strenuous exercise' (OED3), modelled on *warm-up* [1915]. In its turn, *warm-down* is the model for *cool-down* [1976] 'a period of moderate physical activity such as walking or jogging, conducted after vigorous exercise' (OED3), with a shared right component and an opposite left component. The degree of morphosemantic transparency is in both cases 1–1.

Another relevant example of this type, from a V–Adv pattern, is the noun *smoke-easy* [1978] 'a place where cigarettes are smoked illegally; a private smoking club', based on slang *speakeasy* [1889] 'a shop or bar where alcoholic liquor is sold illegally' (OED2). The target, however, is more transparent (1–1) than the model, in which speaking is a consequence (i.e. metonymy) of too much drinking (1–4a).

Lastly, the adjective *hands-on* [1905] 'designating an attitude, policy, etc., characterized by involvement or intervention' (OED3) is obtained from a plural

N–Adv pattern, by analogy with the model *hands-off* [1860] ‘designating an attitude, practice, or policy characterized by a lack of interference or intervention’ (OED3). The latter compounds are analysable as metaphorical (4b–1). Similarly, *brown-out* [1942] ‘a partial black-out’ reproduces the metaphorical pattern of the model *blackout* [1934] (4b–1): i.e. *DARKNESS IS BLACK* like *SEMI-DARKNESS IS BROWN*.

6.3. Compounds with slang or specialised meaning of one or both constituents

Some compounds are increased in opacity because the first, the second, or both constituents have a slang or specialised meaning. The area of slang, both as an informal non-standard variety and as a group-restricted language variety, offers several examples where the constituents have a specific meaning (Mattiello 2008). In other words, slang words are obscured by code mixing. For instance, within the drug community, the compounds *acid jazz* [1988] ‘a genre of dance music incorporating elements of jazz, funk, soul, and hip-hop’ (OED3) and *acid house* [1988] ‘a type of house music characterized by the taking of hallucinogenic drugs’ (OED3) have been coined after the model *acid rock* [1966] (see also *acid head*, *acid trip* [1966]). In both targets and model, the first constituent specifically refers to ‘LSD (lysergic acid diethylamide)’ (degree 3), whereas, for the second constituent, *jazz* and *house* reproduce the polysemy of *rock*, all referring to types of music in these compounds. Therefore, the transparency of these components is blurred by their slang meaning (degree 3).

Specialised (infrequent) meaning also obstructs the understanding of *busgirl* [1914] ‘a girl employed to clear tables in a restaurant’ (OED3), after *busboy* [1904]. Here the shared first constituent *bus* (← *omnibus*) is ambiguous between the standard meaning referring to ‘a large public vehicle’ and the slang meaning of ‘a waiter’s assistant’. The analysis of both target and model is, therefore, 3–1.

With a shared first constituent *hit* we have *hit list* [1976] ‘a list of persons to be assassinated’ (OED2, s.v. *hit*) and *hit squad* [1976] ‘a group of esp. politically-motivated assassins or kidnappers’ (OED2, s.v. *hit*), both after *hit man* [1970] ‘a hired murderer’ (OED2, s.v. *hit*). While the right component is fully transparent (degree 1), the left one is more opaque (degree 3), in that *hit* is polysemic between the standard sense of ‘a blow’ and the slang sense of ‘a killing’ (cf. also the slang sense of ‘a dose of a narcotic drug’ further increasing polysemy).

With a shared second constituent, we have *beefcake* [1949] ‘(a display of) sturdy masculine physique’ (OED2, s.v. *beef*), obtained after *cheesecake* [1929] ‘display of the female body, esp. in photographs’ (OED3). In this case, the model *cheesecake* exhibits metaphorical constituents: i.e. *CAKE IS APPEAL*, *CHEESE IS FEMALE DELICACY* (4a–4a). Metaphorical meaning is

reproduced by *beefcake* (*CAKE IS APPEAL, BEEF IS MALE STRENGTH/VIRILITY*) (4a–4a). In particular, the metaphorical association is here connected, not with the standard meaning of *beef* 'the flesh of an ox or cow, used as food', but with its colloquial sense of ''flesh' (of men); strength, muscular power; effort' (Mattiello 2017b). Another plausible analysis could be metaphorical shift of the whole compound, rather than of each constituent individually (cf. *yellow fever* below).

Another comparable case with a shared first constituent is *dick flick* [2003] 'the testosterone-driven opposite of a 'chick flick'' (*Urban Dictionary*). In this case, the model *chick flick* [1988] 'a film perceived, or marketed, as appealing particularly to women' (OED2, s.v. *chick*) exhibits two polysemic constituents: i.e. *chick* refers to 'a young chicken' in standard language, but to 'a young woman' in slang, and *flick* is slang for 'film', besides its standard meanings of 'a light blow', 'a jerk'. In the target, the polysemy of *flick* is maintained, but the constituent *dick* – a slang word for 'the penis', is not only polysemic, but also metonymic (*MALE ORGAN FOR MALE*). Hence, the degree of opacity of the model (3–3) is increased to 4a–3 in the target. However, since in these compounds both constituents are slang words, they mutually favour (and do not obstruct) slang interpretation.

By contrast, in *kidflick* [1977] 'a cinematographic or video film for children' (OED2, s.v. *kid*), after *kidvid* [1955] 'a television programme or video made for children' (OED2, s.v. *kid*), the shared constituent is not *flick* but *kid*. Moreover, in the target, *flick* is polysemous as well (3–3), whereas *vid* in the model is a colloquial abbreviation of *video*, hence its accessibility is not obstructed semantically but morphologically, by shortening.

A different case is provided by *jungle fever* [2011] 'love for African-American girls' (*Rice University Neologisms Database*), after *yellow fever* [1972], humorous slang for 'strong sexual attraction to people of South-East Asian origin, or to light-skinned African-Americans' (OED3). Here the specialised meaning of both model and target is not connected with the individual constituents, but rather with the whole compounds *yellow fever* and *jungle fever*, which in St.E. respectively refer to 'a severe infectious disease occurring mainly in tropical regions' and 'a form of remittent fever caused by the miasma of a jungle'. However, it is possible to analyse *yellow fever* as 4a–4b, in that *yellow* stands for 'the people who have such a skin colour' and *fever* is a metaphor for 'strong sexual attraction'. A comparable analysis (4a–4b) is in the target *jungle fever*, with shared metaphorical *fever*, and *jungle*, which is a double metonymy for 'Africa', and then for 'African-American girls'.

As for technical jargon, in the analogical compound *daughterboard* [1965] 'a printed circuit board on which are mounted some of the subsidiary components of a microcomputer' (OED3), after *motherboard* [1965] 'a printed

circuit board on which the principal components of a microcomputer are mounted' (OED3), the analysis is 4b–3. Indeed, the non-shared left constituent is metaphorical (*SUBSIDIARY IS DAUGHTER*, *PRINCIPAL IS MOTHER*), while the shared right constituent is specialised, hence restricted in meaning.

Two analogical compounds have been coined in economics after *market share* [1954]: i.e. *stomach share* [1984] 'an informal market share measure used within the food industry' and *wallet share* [1990] 'the proportion of a consumer's disposable income allotted to a single company' (*Wordspy*), both from shortenings (*stomach* (*market*) *share*, *wallet* (*market*) *share*). Here *share* is a polysemic specialised constituent, whereas *stomach* and *wallet* are metonymic (*STOMACH FOR FOOD*, *WALLET FOR INCOME*). Hence, transparency in the targets is decreased compared with the model: from degrees 1–3 to 4a–3.

Another analogical compound or combined form in information technology is *little-endian* [1981] 'designating or relating to computer systems employing byte ordering in which the least significant byte has the lowest address' (OED3), modelled on *big-endian* [1980], the counterpart 'in which the most significant byte has the lowest address', both originally from Jonathan Swift's novel *Gulliver's Travels* (OED). In both target and model, the non-shared left component is polysemic, i.e. metaphorical (*UNIMPORTANT IS SMALL*, *IMPORTANT IS BIG*). The right constituent *endian* is instead fully opaque if taken in isolation and highly specialised in meaning. Indeed, in computing, *endianness* describes the order of byte transmission over a digital link: i.e., in *big-endian* format, the most significant byte is stored first, whereas *little-endian* format stores the least significant byte first. Hence, the overall analysis of target and model is 4b–6.

6.4. Additional examples of compounds with intermediate degrees of opacity/ transparency

There is a group of compounds which exhibits, either in the Variable or in the Invariable Part, an intermediate degree of transparency/opacity.

Degree 3 of the morphosemantic scale of transparency/opacity is, for example, in the left constituent of *brown rice* [1916] 'unpolished rice, with only the husk of the grain removed' (OED2, s.v. *brown*), obtained after *white rice* [1614], but also by analogy with *brown sugar* [1704]. While, in both models, *rice* and *sugar* are fully transparent, *brown* specifically refers to 'unrefined, unpolished', like its opposite *white* refers to 'refined, polished' (cf. *white flour*, *white bread* vs. *brown bread*).

An intermediate degree is also in the right constituent of the compound *inner space* [1958] 'the part of one's mind or personality that is not normally experienced or within one's consciousness' (OED2, s.v. *inner*), after *outer space* [1842], in which the meaning of *space* is restricted to 'an area in one's

mind'. Similarly, in *chairperson* [1971] (OED2), a general word for the models *chairman* [1654] / *chairwoman* [1699] 'the man/woman who presides over a meeting and occupies the seat provided for this function' (OED2), *chair* specifically denotes 'the president's chair', so implying connotations of authority, importance, and power.[8]

An additional example is *Bloody Monday* [1988] 'the first day of the school holidays, on which pupils who have committed an offence may be punished' (*Fifty Years among the New Words*), after *Black Monday* [1735], obsolete slang for 'the first school day after a vacation' (OED3). Here *Monday* specifically refers to the first day of the week in which school starts (degree 3). However, compared with its model, the Variable Part is metonymic (4a, *EFFECT FOR CAUSE*) rather than metaphorical (4b, *BACK TO SCHOOL IS BLACK*, reflecting a negative attitude of lack of enthusiasm and reluctance).

Degree 3 of the scale is also in the right constituent of *dog whisperer* [1998] 'a person who has a natural ability to relate to or connect with dogs' (*Wordspy*), with an Invariable Part *whisperer*, already in *horse whisperer* [1843], with the meaning 'a person who tames or trains an animal typically using body language and gentle vocal encouragement'.

In a similar way, in *narrowcast* [1928] 'to restrict the directions in which a radio transmits, so as to avoid sending signals to areas where listeners are few in number' (OED3), after *broadcast* [1921], the meaning of the verb *cast* 'to throw, to project' is specific (1–3).

6.5. Metonymic compounds

The most consistent group of metonymic analogical compounds exhibits a shared right component with the model. The shared component is typically a body part that stands for 'a person' (Mattiello 2008). Hence, these are exocentric (or *bahuvrihi*) compounds. For instance, there are different compound families in English in which the second constituent *-head* has diverse meanings. In *airhead*[1] [1971] 'a foolish, unintelligent, or frivolous person' (OED3), after *fat-head* [1835], *HEAD STANDS FOR PERSON*, whereas the non-shared component, the noun *air*, metaphorically alludes to 'levity, frivolity, unintelligence', like its corresponding part, the adjective *fat*, in the model (4b–4a) (see § 7.1.2 for differences in word class). A similar case is provided by a slang compound family in which *HEAD STANDS FOR DRUG ADDICT* (Mattiello 2008). These compounds include: *pothead* [1957] 'a habitual user of

[8] The use of *chair* as a converted verb to mean 'to direct (a meeting, etc.) as chairman; to preside over' dates back to 1921 in the OED, thus the compound constituent *chair* is almost certainly a noun, at least in the models.

cannabis' (OED3), *acid head* [1966] 'a person who takes acid' (OED3, s.v. *acid*), *meth head* [1966] 'a habitual user of methamphetamine' (OED3, s.v. *meth*), *crackhead* [1986] 'a person who is addicted to crack cocaine' (OED2, s.v. *crack*), all modelled on the antecedent *dopehead* [1901] and *hophead* [1911] 'drug-addict' (OED2–3). All these analogical compounds could be morphosemantically analysed as 1–4a, in that all non-shared components are co-hyponyms of the superordinate terms *dope* 'drug' or *hop* 'a narcotic drug' of the models. However, transparency is jeopardised (3–4a) by the restricted slang use of the drug names: i.e., *pot* 'cannabis' (from Spanish *potiguaya*), *acid* 'lysergic acid diethylamide', *meth* 'methamphetamine', and *crack* are all polysemous words (see § 6.3).

Still another group of words with a shared, but semantically different *head* component consists of *metalhead* [1982] 'a fan of heavy metal music' (OED3, s.v. *metal*) and *petrolhead* [1980] 'a car enthusiast' (OED3), both based on *breadhead* [1969] 'a person who is motivated by or obsessed with making money' (OED3). Here in the targets we have the same metonymy *HEAD STANDS FOR ENTHUSIAST*, but a different degree of transparency than in the model in the Variable Part. Specifically, the model is analysable as 4b–4a, *bread* being a metaphor for 'material goods', *metalhead* is rather 3–4a, because *metal* is restricted to 'a type of music', and *petrolhead* [1980] is 4a–4a, in that *petrol* metonymically stands for 'car'.

A semantically closely related body part which stands for 'person' is *brain* in *bird brain* [1943] 'a person with a small brain' (OED2, s.v. *bird*), after *beetle-brain* [1593], both analysable as 4b (*SMALL IS BIRD*/*BEETLE*)–4a. By contrast, in *busy brain* [2001] 'a mental state that includes racing thoughts, anxiety, lack of focus, and sleeplessness' (*Wordspy*) and *bypass brain* [2006] 'memory loss and reduced mental functioning after coronary bypass surgery' (*Wordspy*), *BRAIN STANDS FOR MENTAL STATE*. The same metonymy is in the antecedent *milk brain* [1997] 'feelings of disorientation and mental sluggishness reported by some mothers of new-born babies' (*Wordspy*). However, while *milk brain* and *bypass brain* are analysable as metonymic, because *MILK STANDS FOR BREAST-FEEDING* and *BYPASS STANDS FOR CORONARY SURGERY*, the Variable Part in the target *busy brain* is rather metaphorical, in that the adjective *busy* is metaphor for 'confusion, anxiety'. Since all the other first elements preceding *-brain* are nouns, the semantic exception of the first element *busy-* is accompanied by a word-class difference.

The metonymy *MOUTH STANDS FOR TALKING PERSON* is both in the target *motormouth* [1955] 'a person who talks fast and incessantly' (OED3) and in the model *big mouth* [1834], both analysable as 4b–4a, in that *INCESSANTLY IS MOTOR* like *EXCESSIVELY IS BIG*. A comparable example is colloquial *potty mouth* [1969] 'a person who uses obscene language' (OED3),

after *foulmouth* [1692] (both in the *Urban Dictionary*), in which the analysis is again 4b–4a because the Variable Part is a metaphor: i.e. *POTTY*/*FOUL IS BAD LANGUAGE*.

All the above-mentioned examples are different from the case of *computer face* [2010] 'a person's relaxed face after looking at a computer screen for too long' (*Urban Dictionary*), based on *TV face* [2009] 'a person's face after watching TV for too long' (*Urban Dictionary*), where *face* is transparent, while *computer* and *TV* are metonymies (*INSTRUMENT FOR ACTION*). Hence, both target and model are analysable as 4a–1.

The case of the compound family originated after a shared right component *collar* is more complex. Here *collar* is a "double metonymy" (Ruiz de Mendoza & Díez 2002: 512), that is *COLLAR FOR OUTFIT FOR WORK* or *WORKER*. The Variable Part in the family members is instead a metaphor: e.g., *WHITE IS NON-MANUAL, BLUE IS MANUAL, PINK IS FEMALE, GREEN IS CLEAN ENVIRONMENT*, etc. (§ 1). Overall, the analysis of model and targets is 4b–4a. Other possible metonymic patterns in analogical compounds sharing the right component include:

- *PLACE FOR ACTION*: e.g., *air-rage* [1996] 'extreme anger or frustration felt during a flight' (OED3, s.v. *air*), *work rage* [1995] 'extreme workplace anger exhibited by an employee who has been mistreated or fired' (*Wordspy*), and *web rage* [1998] 'extreme anger caused by web frustrations such slow downloads and information that is difficult to find' (*Wordspy*), all after *road rage* [1988] 'violent anger attributed to the stress and frustration of driving a motor vehicle' (OED3) (4a–1);
- *VEHICLE FOR MOTION*: e.g., *carsick* [1908] 'affected with nausea caused by the motion of a car', after *PLACE FOR MOTION* in the models *sea-sick* [*a*1566] 'affected with nausea induced by the motion of a ship at sea' (OED2) and *airsick* [1785] 'sick from the motion of an aircraft' (OED3) (4a–1);
- *ANIMAL FOR ACTION*: e.g., *mouse race* [2003] 'lower-stress life-style' (*Wordspy*), after *rat race* [1937] (§ 3). The Invariable Part *race* is metaphor for 'struggling' (4a–4b);
- *ACTION FOR PERFORMER*: e.g., *try-hard* [1922] 'a person who tries very hard' (OED3) and *work-hard* [1922] 'a person who works very hard' (OED3, s.v. *work*), after *die-hard* [1844] 'one that dies hard' (OED3) (4a–1) (cf. the adjectival meaning of *die-hard* 'extremely committed', converted from the noun).

The case of *first-minute* [n.d.] 'occurring at the earliest possible time' (*WWW*) and *last-second* [1920] 'occurring at the latest possible time before a deadline or

event' (OED3, s.v. *last*) deserves particular attention, in that the two targets share different parts – either the right or the left components – with the model *last-minute* [1908]. However, the metonymy of the right element *MINUTE*/*SECOND FOR VERY SHORT TIME PERIOD* connects both targets with their common model, and the transparency of *first-*/*last-* is also maintained. Later *-second* clearly represents an intensification of earlier *-minute*. The semantic relation of the first element contrasts the most peripheral positions of a time period.

By contrast, in the last group of metonymic analogical compounds, there is a different degree of morphosemantic transparency with respect to the models, in that a transparent constituent is substituted by a metonymic one. These include: *brainwriting* [1913] 'handwriting, regarded as the product of mental characteristics' (OED3), after *handwriting* [1421] (shortened from *brain*(*hand*)*writing*), *granny leave* [2004] 'reduced working hours given to a person who needs to care for an elderly parent' (*Wordspy*), after *maternity leave* [1919], and *cash mob* [2011] 'an event where people support a local retailer by gathering *en masse* to purchase the store's products' (*Wordspy*), after *flash mob* [2003]. The analysis of the targets is, therefore, 4a–1. The metonymies involved are, respectively: *BRAIN FOR MENTAL ACTIVITIES*, *GRANNY FOR OLD PARENT*, and *CASH FOR THE ACTION OF PURCHASING*.

6.6. Metaphorical compounds

A numerous group of metaphorical analogical compounds shares the right element with the model and exhibits a 'colour' as left component. In *orangefield* [2010] 'designating an urban or industrial site that is under-used, but is not contaminated or otherwise unsuitable for development' (*Wordspy*) and *brownfield* [1977] 'designating an (urban) area, which is or has formerly been the site of commercial or industrial activity' (OED2, s.v. *brown*, also in *Wordspy*) the analysis is 4b–1. These compounds share a transparent right constituent and a metaphorical left constituent with the model *greenfield* [1940] 'designating a previously undeveloped site used for commercial development or exploitation' (OED3). Whereas in the model *UNCONTAMINATED IS GREEN*, in the targets *CLEARED AND AVAILABLE FOR REDEVELOPMENT* (said of an industrial site) *IS BROWN*, while *UNDER-USED AND AVAILABLE FOR DEVELOPMENT BUT NOT CONTAMINATED IS ORANGE*. The following example, offered in *Wordspy*, shows that *orange* is viewed as staying in-between *green* and *brown*: "Orange is the new black in the industry. It is a term coined for projects that sit between a greenfield development and a brownfield addition to an existing building." (N. Lenaghan, *Australian Financial Review*, 30/10/2014).

The same morphosemantic analysis 4b–1 is in the targets *gray market*/*grey market* [1934] 'legal but unethical traffic' (OED3) and *white market* [1943] 'authorized dealing in things that are rationed' (OED2, s.v. *white*), both after *black market* [1727]. Here different shades of colour indicate different or no legitimacy: namely, *LEGAL BUT UNETHICAL IS GREY, AUTHORISED IS WHITE* vs. *PROHIBITED IS BLACK*. However, in the chromatic scale with *black* and *white* as the two opposite terms (contradictory colours), in-between *grey* (*market*) unexpectedly precedes *white* (*market*) in the coinage (at least, according to the OED).

Along the same scale, *white money* [2012] 'money obtained by legal means' (OED3, s.v. *white*) and its model *black money* [1939] share the same right component *money*, which is slightly restricted in use (degree 2), and a contradictory, but still metaphorical left component: i.e., in the model, *ILLEGAL IS BLACK*, while in the target *LEGAL IS WHITE* (degree 4b).

The case of *graymail*/*greymail* [1927] 'a mild form of blackmail without demands for money' (OED3), after *blackmail* [1927], is less transparent. The analysis of both target and model is 4b–3. Indeed, the shared right component *mail* is here restricted to its specific sense of 'payment, tax, tribute' (OED3, from Old Icelandic *máli* 'stipulated pay'). The left component, instead, is still metaphorical: namely, in the model, *ILLEGAL WITH EXTORTION IS BLACK*, while in the target *ILLEGAL WITH NO EXTORTION IS GREY*. Thus, *grey* metaphorically represents a milder form of *black*, given the fact that in *greymail* there is no demand for money. Cf. specialised *greenmail* [1983] 'the practice of buying enough shares in a company to threaten a takeover' (OED3), in which *green* is not metaphorical but slang for 'money' (3–3, see § 6.3).

An analogical compound with a colour-related left constituent is *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' (*Rice University Neologisms Database*). Here the model is clearly *Black Friday* [1961] 'the day after Thanksgiving, which traditionally marks the start of the Christmas shopping season' (OED3), in which *black* may metaphorically refer to 'the congestion caused in city centres' (*CONGESTED CITY IS BLACK*, cf. the interpretation as 'a day on which retailers' accounts went from being in the red to being *in the black*' OED3). In the target, instead, the colour metaphor refers to 'the colour of the ribbon that symbolizes breast cancer' (*BREAST CANCER IS PINK*). However, the analysis of both model and target is 4b–1.

Other analogical compounds with a transparent head component shared with the model and a metaphorical non-head component are *hot war* [1947] 'an armed conflict' (OED3, s.v. *hot*), after *cold war* [1945], and *hot warrior* [1950] 'a participant in or advocate of open warfare' (OED3, s.v. *hot*), after *cold-warrior* [1950]. In both cases, the metaphor *ARMED IS HOT* vs. *UNARMED IS*

COLD accompanies a transparent constituent in head position: i.e. 4b–1. Similarly, *soft power* [1990] ‘an approach which avoids coercion and does not rely on military action’ (OED2, s.v. *soft*) and its model *hard power* [1990] ‘power characterized by a coercive approach, often involving military action’ (OED3, s.v. *hard*) exhibit a metaphorical left constituent: *NOT INVOLVING MILITARY ACTION IS SOFT* vs. *INVOLVING MILITARY ACTION IS HARD*.

By contrast, the transparent constituent is in non-head position in the following compounds: *airhead*[2] [1943] ‘an airbase close to an area of active operations where troops can be evacuated by air’ (OED3) and *beachhead* [1940] ‘a fortified position of troops landed on a beach’ (OED2, s.v. *beach*), both after *bridge-head* [1812] (cf. French *tête de pont*). The analysis of both targets and model is 1–4b, the latter degree being represented by the metaphor *FORTIFICATION IS HEAD*. A similar case is *subway desert* [2014] ‘an urban area that is underserved by the city’s subway system’ (*Wordspy*), after *food desert* [1988] ‘a place in which it is difficult to buy food’ (OED3, s.v. *food*), where the triggered metaphor is *ABSENCE IS DESERT*.

The following patterns of analogical compounds, instead, exhibit a shared left constituent with their model:

– *INCESSANTLY IS CHAIN*: *chaindrink* [n.d.] ‘to have drink after drink without pause’ (Kastovsky 1986: 419), after *chainsmoke* [1934] (4b–1);
– *HIGHLY TECHNOLOGICAL IS SMART*: *smartwatch* [1996] ‘a mobile telecommunications device designed to be worn on the wrist, typically with a touch screen display and the ability to connect to a smartphone’ (OED3, s.v. *smart*, also in *Wordspy*) and *smartglasses* [1996] ‘a pair of eyeglasses that includes many of the features of a personal computer’ (*Wordspy*), after *smartphone* [1980] (4b–1);
– *BIRD FEEDER IS CAFETERIA* vs. *BOX IS HOUSE*: *bird cafeteria* [2011] ‘a small box provided for wild birds to feed themselves’ (*Neologisms – New Words in Journalistic Text*), after *bird-house* [1855] (1–4b).

Only in the latter case, the shared constituent is non-metaphorical but transparent.

A more complex case deserves attention in this section. The compound *couch potato* [1979] ‘a person who spends leisure time passively or idly sitting around’ (OED2, s.v. *couch*) has given birth to two target families. One shares the metaphorical right component with the model (*cot potato* [1993], *mouse potato* [1994], *tablet potato* [2010]). The other shares the metonymic left component (*couch tomato* [1988], *couch rat* [1988]), but still has a metaphorical right component. In the first group, the right component *potato* activates the metaphor *IDLE PERSON IS TUBER*, whereas the left component – i.e. a

PLACE FOR ACTION metonymy in the model *couch* – varies. In *cot potato* 'an infant who spends much time watching television' (*Wordspy*), the same metonymy is reproduced (4a–4b). In *mouse potato* 'a person who spends large amounts of leisure time using a computer' (OED3, s.v. *mouse*), a different metonymy is used: i.e. a part-for-whole (*MOUSE FOR COMPUTER*) metonymy (4a–4b). By contrast, in *tablet potato* [2010] 'a person who uses a tablet, especially on a train, airplane, boat, or bus, and is oblivious to everything around him or her' (*Urban Dictionary*), *tablet* specifically refers to 'a small portable computer', therefore giving the interpretation 3–4b. In the second group of analogical compounds, the idea of idleness is metaphorically rendered via different metaphors: namely, *IDLE PERSON IS VEGETABLE* in *couch tomato* [1988] 'a female couch potato' (Algeo 1991: 35) and *IDLE PERSON IS WORTHLESS ANIMAL* in *couch rat* [1988] 'one who spends time watching television' (Algeo 1991: 35). In both analogical compounds, the analysis is 4a–4b. Of course, *couch rat* also reminds of *mall rat* 'a young person who goes to shopping malls to spend time with their friends', but the latter is semantically less close to *couch rat* and its model *couch potato* because of the absent 'idleness' trait.

A novel metaphorical component is finally in *echo boom* [1975] 'a period of marked increase in the birth rate resulting from children born in a previous baby boom themselves becoming parents' (OED3), after *baby boom* [1880] 'a temporary marked increase in the birth rate' (OED3, s.v. *baby*). Here the shared constituent *boom* has a specific meaning connected with 'prosperity, rapid advance, increase' (degree 3), while the non-shared constituent *echo*, which metaphorically alludes to 'repetition, propagation' (degree 4b), substitutes a transparent constituent *baby* (degree 1). Similarly, in *hellseeking* [2011] 'searching for a job in a struggling economy' (*Rice University Neologisms Database*), the metaphor *BAD ECONOMIC SITUATION IS HELL* (degree 4b) substitutes transparent *job* (degree 1) of the model *job-seeking* [1915].

Lastly, a different degree of transparency is in *babymoon* [2015] 'romantic vacation before the baby arrives' (*Urban Dictionary*, also in COCA since 2004) vis-à-vis its model *honeymoon* [1791]. The metaphor *SHORT PERIOD IS MOON* is reproduced in the target, but the metaphor *SWEETNESS IS HONEY* is substituted by a specific use of the noun *baby*, which in the compound *babymoon* refers to 'before having a baby'. Hence, while the model is 4b–4b, the target is 3–4b. By contrast, in *family moon* [1999] 'a honeymoon in which the bride and groom also bring their children from previous marriages' (*Wordspy*) and *weddingmoon* [1995] 'a vacation that includes both a wedding ceremony and a honeymoon' (*Wordspy*), the left component is transparent (1–4b).

Another interpretation of the above-mentioned analogical compounds is that they are formed by discarding the middle element when a new constituent is

added to an existing compound. Thus, *echo* (*baby*) *boom*, *hell* (*job*-)*seeking*, *baby* (*honey*)*moon*, *family* (*honey*)*moon*, and *wedding* (*honey*)*moon* may be their analyses. It is even debatable whether these examples consist of actual compounds, or they are rather analysable as blends (e.g., from *echo* + (*baby*) *boom*, *hell* + (*job*-)*seeking*).

6.7. Opaque compounds

A higher degree of opacity (degree 5) is in the compound adjective *low-rise* [1948] 'of a garment or part of a garment: resting on the hips rather than around the waist' (OED3), opposite to the model *high-rise* [1908]. In both target and model, the meaning of the noun *rise* 'a movement upwards or to a vertical position' is only hinted at, in that the compound adjectives typically apply to trousers, jeans or skirts having a low (or high) waistband.

Another case with an opaque right (but not shared) constituent is *Generation Y* [1992] 'the generation of people following (but in direct contrast to) Generation X' (OED3), which is morphosemantically different from its model *Generation X* [1952] 'a generation of young people about whose future there is uncertainty' (OED3). Here the model *Generation X* is analysable as 1–4b, where *UNCERTAINTY/LACK OF IDENTITY IS X*. In the target, however, *Y* is another variable, and, not by chance, the letter which follows *X* in the alphabet. Therefore, in *Generation Y*, the constituent *Y* is not figurative, nor does it imply uncertainty (though it is, like *X*, a variable), but is specifically related to the fact that this generation follows the previous one. Hence, opacity is increased to 1–5, especially if one does not associate the target to the model.

Even more opaque is the left constituent in *arm candy* [1992] 'a physically attractive (usually female) companion' (OED3, s.v. *arm*), after the metaphorical model *eye candy* [1978] 'an exceptionally attractive person' (OED3, s.v. *eye*). Whereas the model is analysable as 1–4b, i.e. 'a person who is attractive to the eye', the target has an opaque constituent *arm*, which makes the analogical compound analysable as 6–4b. A different case is instead provided by *eye broccoli* [2009] 'an unattractive person' (*Wordspy*), after *eye candy*, in which the metaphorical constituent is opposite in meaning – i.e. *ATTRACTIVE IS CANDY* vs. *UNATTRACTIVE IS BROCCOLI* – but the morphosemantic analysis does not vary (1–4b). The semantic opposition between *candy* and *broccoli*, metaphorically associated with pleasant vs. unpleasant food, is especially evident if one thinks of children, who commonly prefer unhealthy food, such as candies, to healthy vegetables.

A compound family that deserves attention in this section consists of *big*-X words, in which the left constituent is metaphorical (*IMPORTANT IS BIG*). The ancestor in this family is probably *bigwig* [1703] 'a noteworthy or important

person' (OED3), whose right constituent is metonymically related to the large wigs formerly worn by men of distinction or importance (4b–4a). Other family members with the same meaning were coined in the nineteenth century: namely, *big bug* [1826], *big gun* [1834], *big shot* [1861], and *big cheese* [1899]. The latter is still currently used in this sense, as attested in the *Urban Dictionary* [2006]. All the family members are denotative (but not connotative) synonyms, their connotative features being connected with the person they are referred to (e.g. businessmen, politicians, sportsmen, etc.).

A more recent compound in this family is *big wheel* [1942] 'an important person' (OED3, s.v. *big*) (see also *big fish* [1827], § 3). Whereas some of the right constituents in this family may have a metaphorical explanation (e.g. *fish* may be referred to 'a person whom it is desirable to catch or hook'), others are fully opaque, especially, *bug* (having negative connotations), *cheese*, and *wheel*. The latter examples belong to degree 6 of the morphosemantic scale.

7. Quantitative results and discussion

This section is devoted to quantitative results on analogical compounds and their constituents vis-à-vis their models. The overall number of analogical compounds selected for the analysis is 115 (see Appendix for the entire list of targets, models, and constituents' (morpho)semantic/syntactic analysis). One of them (*brown rice*) has two models (i.e. *brown sugar* and *white rice*), with which it respectively shares the same first and second component. Only two of them are three-member compounds (*second-hand drinking*, *third-hand smoke*), whose complex left members (*second-hand*/*third-hand*) form themselves subordinate compounds. There is no substitution of a simplex element with a compound (or vice versa) in our data, with the partial exception of *bypass* (vs. *milk*) *brain*, in which the components of *bypass* are more strictly bonded than in other compounds. This shows the tendency of analogical compounds to reproduce the binary relation of their models, which is the most "natural" (hence, the preferred) relation in universal naturalness parameters (Dressler *et al.* 1987).

This binary preference is also shown by many originally three-member compounds where the intermediate member drops: e.g., *baby*-(*honey*)-*moon*, *brain*-(*hand*)-*writing*, *echo*-(*baby*)-*boom*, *family*-(*honey*)-*moon*, *hell*-(*job*)-*seeking*, *kite*-(*surf*)-*board*, *stomach*-(*market*)-*share*, etc. In these cases, it is expectable that the morphosemantic analyses of the targets and those of the models do not correspond, nor do the semantic relations between the compounds' components. Morphologically, these could be analysed as blends merging two words, one of which is a compound (← *baby* + (*honey*)*moon*, ← *brain* + (*hand*)*writing*, etc.). Morphosemantically, the head is partially obscured by shortening: i.e., a *babymoon* is 'a honeymoon before having a

baby', *brainwriting* is 'handwriting involving the mental activities of the brain', and so on. It is therefore debatable whether these novel complex words could be viewed as products of substitution by analogy, or rather as shortenings.

As for hyphenation, the analogical compounds in our dataset are variously spelt as two separate words (52 instances), as one word (34), or as hyphenated (29). With the exception of 18 cases, most of them tend to maintain the same spelling as their models.

7.1. Variable Part

By definition, analogical (target) compounds are made up of an Invariable Part, which is shared with the model, and a Variable Part. While the importance of a shared element is essential in the identification of novel analogical compounds, the significance of the non-shared Variable Part should also be taken into consideration. In particular, Table 2 shows the importance of maintaining the same transparency/opacity degree in the Variable Part, as well as the same part-of-speech, and semantic relation with the Invariable Part with respect to the model.

Table 2. Analogical (target) compound constituents and their relation with the model constituents

Target (vs. Model)	Variable Part	Transparency/Opacity Degree			Part-of-Speech		Semantic Relation with Invariable Part	
		Same	Increased	Decreased	Same	Different	Same	Different
First Constituent	91	70	11	10	84	7	85	30
Second Constituent	25	23	2	0	25	0		
Total	116	93	13	10	109	7	85	30

7.1.1. Transparency/opacity degree

Among the analogical compounds in our dataset, 78.4% (91) exhibit a non-shared First Constituent, and only 21.5% (25) a non-shared Second Constituent. The case of *brown rice* counts as two Variable Parts, one in first and one in second position. Therefore, the Variable Part tends to be in non-head position (on the left), whereas the Invariable Part is generally in head position (on the right). In the three-member compound *third-hand smoke*, only the left-most component varies (cf. *second-hand smoke*). In converted compounds from phrases, the verb varies whereas the preposition or the adverb on the right is kept.

Table 2 also shows that, in the Variable Part of analogical compounds, the transparency/opacity degree tends to be maintained in 93 cases (80.1%), or better increased (11.2%) than decreased (8.6%). In particular, optimal transparency (degree 1) of the Variable Part is reproduced in the first constituent (27 cases) or in the second constituent (15 cases). Figurative meaning of the Variable Part also tends to be preserved in the targets (cf. Mattiello 2017b), especially metaphorical meaning (22 cases in the first constituent and 4 cases in the second constituent) and metonymic meaning (14 cases in the first constituent and 1 case in the second constituent). Finally, semantic specification (degree 3), as in slang or specialised meaning, is also reproduced in the Variable Part (6 cases in the first constituent and 3 cases in the second constituent). The Variable Part is rarely opaque (degrees 5-6) in our data, and when it is (only in *arm candy*, *eternity leave*), the motivation seems to be jocularity or rhetorical effect.

This equivalence in transparency degree is made possible thanks to the semantic similarity between the Variable Parts in targets and models. As anticipated in § 3, the primary relations linking the Variable Parts are:

– Near identity, quasi-synonymy (e.g. *hop–dope*, *mouse–rat*), or even true synonymy (e.g. *big–great*, *call–ring*);
– Contradictory opposition: e.g., *cool–warm*, *down–up*, *first–last*, *hot–cold*, *little–big*, *low–high*, *narrow–broad*, *slow–fast*, *small–big*, *soft–hard*, *white–black*;
– Contrary opposition: e.g., *daughter–mother*, *girl–boy*, *father–mother*, *inner–outer*, *on–off*;
– Other contrasts: e.g., co-hyponymy (*air–road*, *bar–restaurant*, *be–fish–love–read–sleep–sit*, *beef–cheese*, *bird–beetle*, *black–blue–brown–green–grey–orange–pink–white*, *broccoli–candy*, *cafeteria–house*, *computer–TV*, *cot–couch*, *dog–horse*, *drink(ing)–smok(ing)*, *ear–eye*, *fax–ring*, *flick–vid*, *glasses–watch–phone*, *house–jazz–rock*, *listen–read*, *look(ie)–talk(ie)*, *moon–earth*, *rice–sugar–dairy*, *sand–snow*, *smoke–speak*, *third–second*, *tomato–potato*, *try–work–die*, *walk–drive*, *Y–X*), hyponym–hypernym (e.g. *acid*/*crack*/*meth*/*pot–dope*), hypernym–hyponym (*person–man*/*woman*), or meronymy (*second–minute*, *man–squad*);
– Less evident semantic relations: e.g., 'by *car*' and 'by *sea*' are two ways of travelling, *dick* and *chick* are two slang words, the former referring to 'the male organ' and the latter to 'a young woman'. By contrast, *tablet* (in *tablet potato*) is perhaps more closely related to *mouse*, in the intermediate model *mouse potato*, than to *couch* in *couch potato*.

Other similarities can be identified between the Variable Parts in targets and models, namely phonological, through alliteration and perfect or imperfect rhymes (*beach–bridge*, *bloody–black*, *cash–flash*, *eternity–maternity*, *wheel–wig*), and morphotactic (*handie–walkie*), or both (*catching–casting*) (see Mattiello 2016 for similarity scales).

When the Variable Parts in targets and models do not exhibit the same degree of transparency, there may be a shift towards more transparency, as in *blame* (1) vs. *brain* (4a)-*storming*, or *smoke* (1) vs. *speak* (4a)-*easy*. However, the opposite tendency is also possible: e.g., *big wig* (6) vs. *wheel* (4a), *granny* (4a) vs. *maternity* (1) *leave*, and *arm* (6) vs. *eye* (1) *candy*, in spite of the co-hyponymy between the two body parts.

7.1.2. Part-of-speech

The part-of-speech of the Variable Part is also preferably shared between target and model compounds, with rare exceptions (6%, 7 examples). When it varies, it is often a syntactic change, rather than a functional one: e.g., in *motormouth* (after *big mouth*) and *potty mouth* (after *foulmouth*), *motor* and *potty* are nouns with a modifying function, which is close to that of attributive adjectives.

It is interesting to remark that, even when the Variable Parts are completely unrelated from the semantic viewpoint, their word class is nevertheless maintained. This is the case with *subway desert* ← *food desert*, *mouse potato* ← *couch potato*, and *couch rat* ← *couch potato*, in which all Variable Parts are nouns. This is not coincidentally, since nouns constitute the largest number of English vocabulary items. Moreover, English has a general noun bias (Tardif, Gelman & Xu 1999).

By contrast, when adverbs or particles are maintained in the targets (e.g. *warm-down*, *hands-on*), their relevance is higher because of their relative infrequency as compound constituents. The role of inflectional markers maintained in targets is also central: e.g., the Saxon genitive added to the Variable Part in *Father's Day*, after *Mother's Day*. In *metalhead* and *petrolhead*, both after *breadhead*, the Variable Parts *metal*, *petrol*, and *bread* also share their uncountable nature.

7.1.3. Semantic relation with Invariable Part

The semantic relation between the Variable and Invariable constituents is another variant linking the targets to (or distinguishing them from) their models.

In 26.9% (31) of the examples in our database (20% if we exclude originally three-member compounds), the semantic relation between the constituents changes. This confirms that there is often no correlation between the constituent

syntactic pattern (cf. 6% in § 7.1.2) and their semantic relation. There is a closer correlation, instead, between the different degree of morphosemantic transparency of the Variable Part and its semantic link to the Invariable Part. In other words, the relation between Variable and Invariable Parts changes if the Variable Part increases or decreases in transparency compared with the model. The reverse is not always true. For instance, both *carsick* and *seasick* are analysable as 4a–1, in spite of the different semantic relation between the constituents: i.e., 'sick because of the motion of a car' vs. 'sick because of the motion of a ship at sea'. An equivalent relation would have occurred, instead, between the target *carsick* and the inexistent model *shipsick*, or between the potential target *roadsick* and the model *seasick*.

When figurative (i.e. metaphorical or metonymic) language intervenes, it is commonly reproduced in the target. However, the relation between the constituents may vary: e.g., in *jungle fever* and *yellow fever*, the metaphor *FEVER IS SEXUAL ATTRACTION* is reproduced, but the different metonymies change the constituents' relation, namely: 'sexual attraction to people whose origin is from the African jungle' vs. 'sexual attraction to people having a yellow skin'.

Even when the compounds are fully transparent, such as *puppy leave* and its model *maternity leave*, the relation between the elements may be different: i.e. 'leave from work to take care of a new puppy' vs. 'leave from work in the weeks before and after giving birth (maternity)'. By contrast, the compound *paternity leave* [1973] 'a short period of authorized absence from employment granted to a father after or shortly before the birth of his child' (OED3, s.v. *paternity*) – with an opposite Variable Part, but not in our database – also reproduces the semantic relation that we find in *maternity leave*. On the other hand, *eternity leave* and *granny leave*, from the same model, are neither fully transparent nor reproduce the same semantic relation.

The semantic relation between the compound constituents is even independent of the semantic similarity between the Variable Parts in target and model compounds. For instance, we would expect the same semantic relation between the components of *greenmail* and *blackmail*, or between the elements making up *Pink Friday* and *Black Friday*, especially because of the co-hyponymy between *green* and *black* or *pink* and *black*. However, the specialised (slang) meaning of *green* 'money' and the different metaphor activated by *pink*, symbolising 'breast cancer', influence the overall interpretation of the novel compounds.

8. General discussion on compound families

The formation of compound families heavily relies on the analogical process. It is by substituting one of the elements in a compound – either the right (head)

constituent or, more frequently, the left (non-head) constituent – that we can obtain an analogical set forming a compound family.

The role played by the Invariable Part and the Variable Part in this process is different, but equally fundamental. In particular, the Invariable Part is the shared element among the family members, and it is often the most salient member in the compounds, either from the semantic viewpoint (e.g. the element with a specific (specialised or slang) meaning), or from the syntactic viewpoint (e.g. the element with the less frequent word class as compound constituent), or both. On the other hand, the Variable Part is the (paradigmatic) substitute in the analogical proportion, and it is often semantically, but also phonologically and morphotactically similar to the substituted element in the model compound.

Hence, while the Invariable Part establishes the model compound family for a novel analogical compound, the Variable Part determines the difference between nuclear and extended compound family. Indeed, in a nuclear compound family, the Variable Parts in all members: 1) are semantically similar, i.e. related by "a distinct semantic link […] (typically opposition, co-hyponymy, synonymy)" (Klégr & Čermák 2010: 236), 2) are syntactically related (i.e. they belong to the same part-of-speech), and 3) exhibit the same syntagmatic relation with the other compound constituent.

As observed (§ 7.1.1), one of the privileged semantic links shared by the Variable Parts in a nuclear compound family is co-hyponymy (followed by contradictory/contrary opposition and, with lower frequency, (near-)identity and meronymy). Co-hyponymy leads to the rapid lexical expansion of a compound family, in that it allows us to create novel family members by substituting the Variable Part with one of its co-hyponyms. The higher the number of available co-hyponyms, the more productive the pattern of the compound family. For instance, the family ending in *-collar* (from *white-collar*) has become productive because the Variable Part (*white*) is a subordinate of the superordinate term 'colours'. Originally, *blue* substituted quasi-antonymic *white*, but later a range of other colours became variables (*pink*, *green*, *black*, etc.), and others are potential variables (e.g. *yellow* or *red*). All these Variable Parts display metaphorical meaning, thus sharing the same transparency/opacity degree (4b) along our scale. Similarly, the productivity of the compound family with *-in* as second constituent (from *sit-in*) is linked to the existence of co-hyponym verbs (*be*, *fish*, *love*, *read*, *sleep*) denoting actions or states which substitute *sit* in the analogical proportion. In the latter example, the productivity of the *-in* family has contributed to the conversion of *in* from a preposition to a suffix (OED).

A nuclear compound family is the basic subfamily of an extended compound family, accompanied by other, semantically different and mostly derived subfamilies. This holds also for the Invariable Part of compound families. For

example, the family of X-*head* compounds divides into a nuclear subfamily with *head* in its original meaning, as in *forehead, back-head, big-head* 'disease of livestock characterized by swelling of the head', the metonymically derived subfamily where *head* signifies *PERSON*, as in the examples of § 6.5, and another subfamily where it designs inanimate salient, peripheral objects, such as *bedhead, dog-head, figurehead,* etc. Thus, a compound family divides orthogonally in subfamilies semantically defined by either the Variable or the Invariable compound constituent. What is important for our topic is that analogical compounds are created within the same subfamily.

The part-of-speech of the Variable Parts in a nuclear family is also shared (e.g. *blue, green, pink* are adjectives; *be, read, sleep* are verbs, etc.). However, this similarity in word class is of fundamental importance especially with less frequent syntactic patterns, such as V–Adv (*try-hard, work-hard, die-hard,* or *smoke-easy, speak-easy*), or V–Prep (*cool-down, warm-down, warm-up*). While the diachronic study of analogical words can help distinguish between model and target forms, a shared infrequent syntactic pattern can help the association between them. For instance, a V–*hard* pattern was activated for new compounds meaning 'a person who Vs hard', while a V-*easy* pattern was used to interpret novel compounds meaning 'a place where we (literally or metaphorically) V easy'. On the other hand, when the syntactic pattern is very frequent, such as N–N or A–N, its significance in the association of targets to models decreases.

Finally, another important aspect for the formation of a compound family is the semantic relation that links the compound constituents. In section 6.6, we discussed the case of the compound families originated from the same model *couch potato*. The family X-*potato* consists of *cot potato, mouse potato,* and *tablet potato,* whereas the family *couch*-X consists of *couch tomato* and *couch rat*. Although the meaning of all these forms is comparable – i.e. they all refer to idle people spending much time in passive behaviours – there is a subtler distinction among them based on the semantic relation between the components. Thus, *cot potato* and *couch potato* are semantically closer because they refer to 'people who spend leisure time sleeping on a cot/sitting on a couch'. On the other hand, within the same family, *mouse potato* is probably a more precise model for *tablet potato,* being they more specifically connected with the use of computers: 'a person who spends his/her time using (the mouse of) a computer/a small portable computer'. In the other family, including *couch potato, couch tomato,* and *couch rat,* the relation between the constituents is again one connected with the action of 'sitting on a couch', while the Variable Part metaphorically alludes to 'a lazy person'.

Our orthogonal subdivision of compound families into subfamilies challenges the current psycholinguistic practice of defining Family Size simply by adding up all compounds with the same constituent in the same position

(Bertram, Baayen & Schreuder 2000; Mulder *et al.* 2014). Semantic differences between subfamilies can make a compound family very heterogeneous, and our finding that analogies are produced within the same subfamily indicates that similarity effects (also called "gang" effects) should be stronger within the same subfamily than across different subfamilies of the same family. Since there are also semantic similarities between subfamilies of different compound families, for example between the two subfamilies X-*head* and X-*brain*, where -*head* and -*brain* both design persons, it would be interesting to test experimentally whether (and under which conditions) similarity effects are higher between semantically different subfamilies of the same compound family or between semantically similar subfamilies of different compound families.

9. Conclusions

The research questions that we posed in the Introduction to this study concerned the degree of morphosemantic transparency/opacity of novel analogical compounds and its role in the formation of compound families. This study has demonstrated that novel analogical compounds, although formed by productive compounding rules, are by far not predominantly fully transparent. Their degrees of transparency/opacity are based on their models within the same word-formation family. In other words, within the domain of a productive word-formation rule, constituent families are crucial for attracting new family members. Novel analogical compounds tend to maintain the same transparency/opacity degree as their models in their Variable Part (80.1% in our dataset), or, at least, they are not more transparent than their models. The morphosemantic transparency (degree 1) of the Variable Part in fully transparent endocentric compounds is entirely reproduced, and so is the figurative meaning of metonymic (degree 4a) and metaphorical (degree 4b) compounds, with very few exceptions (e.g. *big wheel* 6 ← *bigwig* 4a, *greenmail* 3 ← *blackmail* 4b, *smoke-easy* 1 ← *speakeasy* 4a). Specialised (especially slang) meaning also tends to be maintained in target compounds, as in *pothead* ← *dopehead*, with a degree 3 replicated in the Variable Part.

By contrast, when the degree of transparency/opacity is not replicated, there may be a tendency towards more transparency (11.2%, e.g. *blamestorming* 1 ← *brainstorming* 4a). The reverse process (towards a higher degree of opacity) may be a consequence of the deletion of the intermediate constituent in originally three-member compounds (e.g. *brainwriting* 4a ← *brain* + (*hand*)*writing* 1). The deletion process, however, may also obtain novel compounds whose opacity degree is decreased compared with the model (e.g. *family moon* 1 ← *family* + (*honey*)*moon* 4b). Differences at the morphosemantic level are often counterbalanced by similarities at the phonological level

(*dick*/*chick flick*, *smoke-*/*speakeasy*). This is also valid for compounds with higher degrees of opacity, such as *eternity leave* 5 ← *maternity leave* 1.

Therefore, both Invariable Part and Variable Part play a key role in the interpretation of novel analogical compounds and in the identification of their model, or model compound family. The Invariable Part constitutes the shared element, hence the part which remains constant, both syntactically and (morpho)semantically. The Variable Part, instead, varies, but, in any case, it keeps a distinct semantic link with the Variable Part in the model(s). This semantic link (be it near-synonymy, opposition, or co-hyponymy) establishes the nuclear compound family members, distinguishing them from the extended family members.

In nuclear families and subfamilies of compounds, the part-of-speech of the constituents, their degree of transparency/opacity, and their semantic relation are reproduced in all members of the analogical set, both old and novel ones.

REFERENCES

Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 425-440. Oxford: Blackwell.

Arndt-Lappe, Sabine. 2015. Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation - An International Handbook of the Languages of Europe: Volume 2*, 822-841. Berlin: De Gruyter Mouton.

Arndt-Lappe, Sabine & Melanie J. Bell. Under revision. Analogy and the nature of linguistic generalisation: Locality, generality, and variability in English compound stress. Manuscript, Heinrich-Heine-Universität Düsseldorf & Anglia Ruskin University.

Aronoff, Mark. 1976. *Word formation in Generative Grammar*. Cambridge, MA: The MIT Press.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bauer, Laurie & Paul Nation. 1993. Word families. *International Journal of Lexicography* 6(4). 253-279.

Bell, Melanie J. 2012. The English noun-noun construct: A morphological and syntactic object. In Angela Ralli, Geert Booij, Sergio Scalise & Athanasios Karasimos (eds.), *Morphology and the architecture of grammar. Proceedings of the 8th International Morphology Meeting*, 59-91. Patras: University of Patras; http://morbo.lingue.unibo.it/mmm.

Bell, Melanie J. & Ingo Plag. 2013. Informativity and analogy in English compound stress. *Word Structure* 6(2). 129-155.

Bell, Melanie J. & Martin Schäfer. 2016. Modelling semantic transparency. *Morphology* 26(2). 157-199.

Benczes, Réka. 2006. *Creative compounding in English*. Amsterdam/Philadelphia: Benjamins.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150-177.

Bertram, Raymond, R. Harold Baayen & Robert Schreuder. 2000. Effects of family size for complex words. *Journal of Memory and Language* 42. 390-405.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt and Company.

Booij, Geert. E. 2010. Compound constructions: Schemas or analogy? A construction morphology perspective. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 93-108. Amsterdam: John Benjamins.

Bourque, Stephen Yves. 2014. *Toward a typology of semantic transparency: The case of French compounds*. Unpublished Ph.D. Thesis. Department of French Studies, University of Toronto.

Brewer, Charlotte. 2014. *Treasure-house of the Language: The Living OED*. New Haven, CT/London: Yale University Press.

Bybee, Joan L. 1988. Morphology as lexical organization. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology. Approaches in modern linguistics*, 119-141. San Diego: Academic Press.

Chomsky, Noam. 1957. *Syntactic structure*. The Hague: Mouton.

De Jong, Nivja H., Laurie B. Feldman, Robert Schreuder, Matthew Pastizzo & Baayen R. Harald. 2002. The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language* 81(1). 555-567.

Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. In Ursula Doleschal & Anna M. Thornton (eds.), *Extragrammatical and Marginal Morphology*, 1-10. München: Lincom Europa.

Dressler, Wolfgang U., Willi Mayerthaler, Oswald Panagl & Wolfgang U. Wurzel. 1987. *Leitmotifs in Natural Morphology*. Amsterdam/Philadelphia: Benjamins.

Dressler, Wolfgang U. & Barbara Tumfart. 2017. New corpus-linguistic approaches to the investigation of poetic occasionalisms: The case of Johann Nepomuk Nestroy. *Yearbook of the Poznan Linguistic meeting* 3. 155-166.

Fabb, Nigel. 1998. Compounding. In Andrew Spencer & Arnold Zwicky (eds.), *The handbook of morphology*, 66-83. Oxford: Blackwell.

Fertig, David L. 2013. *Analogy and morphological change*. Edinburgh: Edinburgh University Press.

Gagné, Christina L. 2009. Psycholinguistic perspectives. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 255-271. Oxford: Oxford University Press.

Gagné, Christina L. & Thomas L. Spalding. 2014. Typing time as an index of morphological and semantic effects during English compound processing. *Lingue e Linguaggio* 13(2). 241-262.

Gardani, Francesco. 2013. *Dynamics of morphological creativity*. Leiden: Brill.

Goswami, Usha. 1991. Analogical reasoning: What develops? A review of research and theory. *Child Development* 62(1). 1-22.

Hay, Jennifer & Harald Baayen. 2002. Parsing and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2001*, 203-235. Dordrecht: Kluwer Academic Publishers.

Hockett, Charles F. 1968. *The state of the art*. The Hague/Paris: Mouton.

Jarema, Gonia, Céline Busson, Rossitza Nikolova, Kyrana Tsapkini & Gary Libben. 1999. Processing compounds: A cross-linguistic study. *Brain and Language* 68. 362-369.

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

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

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

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

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.

Levi, Judith N. 1978. The syntax and semantics of complex nominals. New York: Academic Press.

Libben, Gary. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language* 61. 30-44.

Libben, Gary. 2008. Compounding and cognition. *ICCS Proceedings*, 27-29 July 2008. Seoul, Korea.

Libben, Gary. 2010. Compound words, semantic transparency, and morphological transcendence. In Susan Olsen (ed.), *New impulses in word-formation*, 317-330. Hamburg: Buske.

Libben, Gary & Silke Weber. 2014. Semantic transparency, compounding, and the nature of independent variables. In Franz Rainer, Francesco Gardani, Hans Christian Luschützky & Wolfgang U. Dressler (eds.), *Morphology and meaning*, 205-221. Amsterdam/Philadelphia: Benjamins.

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

Liu, Dilin & Hongwei Zhan. 2015. The use of the *-free* compound and *free of* and *free from* phrasal constructions: A diachronic and synchronic study. *Journal of English Linguistics* 43(3): 201-226.

Marle, Jaap van. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Mattiello, Elisa. 2008. *An introduction to English slang: A description of its morphology, semantics and sociology*. Monza: Polimetrica.

Mattiello, Elisa. 2013. *Extra-grammatical morphology in English. Abbreviations, blends, reduplicatives, and related phenomena*. Berlin/Boston: De Gruyter Mouton.

Mattiello, Elisa. 2016. Analogical neologisms in English. *Italian Journal of Linguistics* / *Rivista di Linguistica* 28(2). 103-142.

Mattiello, Elisa. 2017a. *Analogy in word-formation. A study of English neologisms and occasionalisms*. Berlin/Boston: De Gruyter Mouton.

Mattiello, Elisa. 2017b. The impact of figuration on word-formation: The role of figurative language in the production and interpretation of novel analogical compounds. *Textus* XXX(1). 17-32.

Motsch, Wolfgang. 1981. Der kreative Aspekt in der Wortbildung. In Leonhard Lipka (ed.), *Wortbildung*, 94-118. Darmstadt: Wissenschaftliche Buchgesellschaft.

Mugglestone, Lynda, 2005. *Lost for words: The hidden history of the Oxford English Dictionary*. New Haven, CT/London: Yale University Press.

Mulder, Kimberley, Ton Dijkstra, Robert Schreuder & R. Harald Baayen. 2014. Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language* 72. 59-84.

Partee, Barbara Hall, Alice G. B. ter Meulen & Robert Eugene Wall. 1990. *Mathematical methods in linguistics*. Dordrecht: Kluwer Academic.

Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer.

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

Plag, Ingo. 2010. Compound stress assignment by analogy: The constituent family bias. *Zeitschrift für Sprachwissenschaft* 29(2). 243-282.

Pollatsek, Alexander & Jukka Hyönä. 2005. The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes* 20(1/2). 261-290.

Rainer, Franz, Wolfgang U. Dressler, Francesco Gardani & Hans Christian Luschützky. 2014. Morphology and meaning: An overview. In Franz Rainer, Francesco Gardani, Hans Christian Luschützky & Wolfgang U. Dressler (eds.), *Morphology and meaning*, 3-46. Amsterdam/Philadelphia: Benjamins.

Ruiz de Mendoza Ibáñez, Francisco José & Olga Isabel Díez Velasco. 2002. Patterns of conceptual interaction. In René Dirven & Ralph Pörings (eds.), *Metaphor and metonymy in comparison and contrast*, 489-532. Berlin/New York: Mouton de Gruyter.

Scalise, Sergio. 1992. Compounding in Italian. *Italian Journal of Linguistics / Rivista di Linguistica* 4(1). 175-199.

Schäfer, Martin. 2018. *The semantic transparency of English compound nouns*. Berlin: Language Science Press.

Schmidtke, Daniel, Victor Kuperman, Christina C. Gagné & Thomas L. Spalding. 2016. Competition between conceptual relations affects compound recognition: The role of entropy. *Psychonomic Bulletin & Review* 23(2). 556-570.

Schwaiger, Sonja, Jutta Ransmayr, Katharina Korecky-Kröll, Sabine Sommer-Lolei & Wolfgang U. Dressler. 2017. Scaling morphosemantic transparency/opacity: A corpus-linguistic and acquisitionist study of German diminutives. *Yearbook of the Poznań Linguistic Meeting* 3(1). 141-153.

Skousen, Royal & Thereon Stanford. 2007. *AM: Parallel*. Provo, UT: Brigham Young University.

Smith, Viktor, Daniel Barratt & Jordan Zlatev. 2014. Unpacking noun-noun compounds: Interpreting novel and conventional food names in isolation and on food labels. *Cognitive Linguistics* 25(1). 99-147.

Talamo, Luigi, Chiara Celata & Pier Marco Bertinetto. 2016. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure* 9(1). 72-102.

Tardif, Twila, Susan A. Gelman & Fan Xu. 1999. Putting the 'noun bias' in context: A comparison of English and Mandarin. *Child Development* 70(3). 620-635.

DATA SOURCES

Algeo, John (ed.). 1991. *Fifty years among the new words: A dictionary of neologisms, 1941-1991*. Cambridge: Cambridge University Press.

Corpus of contemporary American English. 1990-2017. COCA. Brigham Young University.

Available at http://corpus.byu.edu/coca/ (27 January, 2018).

Neologisms – new words in journalistic text. 1997-2012. Birmingham City University. Available at http://rdues.bcu.ac.uk/neologisms.shtml (11 December, 2016).

Oxford English dictionary online. 1989-2018. 2nd edn./3rd edn. (OED2/OED3). Oxford: Oxford University Press. Available at http://www.oed.com/ (11 December, 2016).

The Rice University neologisms database. 2004-2014. Rice University. Available at http://neologisms.rice.edu/ (11 December, 2016).

Urban dictionary. 1999-2017. Available at http://www.urbandictionary.com/ (11 December, 2016).

Wordspy – The word lover's guide to new words. 1995-2017. Available at http://www.wordspy.com/ (11 December, 2016).

Appendix: Analogical (target) compounds and their models: Analysis of the components' transparency/opacity degree, part-of-speech, and semantic relation.

Target	Model	Morphosemantic Analysis		Part-of-Speech		Semantic Relation
		Target	Model	Target	Model	
acid head	dopehead, hophead	3–4a	3–4a	N–N	N–N	=
acid house	acid rock	3–3	3–3	N–N	N–N	=
acid jazz	acid rock	3–3	3–3	N–N	N–N	=
airhead[1]	fat-head	4b–4a	4b–4a	N–N	A–N	=
airhead[2]	bridge-head	1–4b	1–4b	N–N	N–N	=
air-rage	road rage	4a–1	4a–1	N–N	N–N	=
arm candy	eye candy	6–4b	1–4b	N–N	N–N	≠
babymoon	honeymoon	3–4b	4b–4b	N–N	N–N	≠
beachhead	bridge-head	1–4b	1–4b	N–N	N–N	=
beefcake	cheesecake	4a–4a	4a–4a	N–N	N–N	=
be-in	sit-in	1–5	1–5	V–Prep	V–Prep	=
big gun	great gun	4b–6	4b–6	A–N	A–N	=
big wheel	bigwig	4b–6	4b–4a	A–N	A–N	=
bird brain	beetle-brain	4b–4a	4b–4a	N–N	N–N	=
bird cafeteria	bird-house	1–4b	1–4b	N–N	N–N	=
black-collar	white-collar	4b–4a	4b–4a	A–N	A–N	=
blamestorming	brainstorming	1–4b	4a–4b	N–N	N–N	≠
Bloody Monday	Black Monday	4a–3	4b–3	A–N	A–N	≠
blue-collar	white-collar	4b–4a	4b–4a	A–N	A–N	=
brainwriting	handwriting	4a–1	1–1	N–N	N–N	≠
brown rice	white rice/ brown sugar	3–1	3–1	A–N	A–N	=
brownfield	greenfield	4b–1	4b–1	A–N	A–N	=
brown-out	blackout	4b–1	4b–1	V–Prep	V–Prep	=
busgirl	busboy	3–1	3–1	N–N	N–N	=
busy brain	milk brain	4b–4a	4a–4a	A–N	N–N	≠
bypass brain	milk brain	4a–4a	4a–4a	N–N	N–N	≠
café-bar	café-restaurant	1–1	1–1	N–N	N–N	=

callback	ringback	1–1	1–1	V–Prep	V–Prep	=
carsick	sea-sick, airsick	4a–1	4a–1	N–N	N–N	≠
cash mob	flash mob	4a–1	4a–1	N–N	N–N	≠
chaindrink	chainsmoke	4b–1	4b–1	N–V	N–V	=
chairperson	chairman/woman	3–1	3–1	N–N	N–N	=
computer face	TV face	4a–1	4a–1	N–N	N–N	=
cool-down	warm-down	1–1	1–1	V–Prep	V–Prep	=
cot potato	couch potato	4a–4b	4a–4b	N–N	N–N	=
couch rat	couch potato	4a–4b	4a–4b	N–N	N–N	=
couch tomato	couch potato	4a–4b	4a–4b	N–N	N–N	=
crackhead	dopehead	3–4a	3–4a	N–N	N–N	=
dairy-free	sugar-free, gluten-free	1–2	1–2	N–A	N–A	=
daughterboard	motherboard	4b–3	4b–3	N–N	N–N	=
dick flick	chick flick	4a–3	3–3	N–N	N–N	≠
dog whisperer	horse whisperer	1–3	1–3	N–N	N–N	=
ear-witness	eyewitness	1–1	1–1	N–N	N–N	=
echo boom	baby boom	4b–3	1–3	N–N	N–N	≠
eternity leave	maternity leave	5–1	1–1	N–N	N–N	≠
eye broccoli	eye candy	1–4b	1–4b	N–N	N–N	=
family moon	honeymoon	1–4b	5b–4b	N–N	N–N	≠
Father's Day	Mother's Day	1–1	1–1	N's–N	N's–N	=
father-substitute	mother-substitute	1–1	1–1	N–N	N–N	=
fax-back	ringback	1–1	1–1	V–Prep	V–Prep	=
first-minute	last-minute	1–4a	1–4a	A–N	A–N	=
fish-in	sit-in	1–5	1–5	V–Prep	V–Prep	=
Generation Y	Generation X	1–5	1–4b	N–N	N–N	=
granny leave	maternity leave	4a–1	1–1	N–N	N–N	≠
green-collar	white-collar	4b–4a	4b–4a	A–N	A–N	=
grey market	black market	4b–1	4b–1	A–N	A–N	=
greenmail	blackmail	3–3	4b–3	N–N	A–N	≠
greymail	blackmail	4b–3	4b–3	A–N	A–N	=
hands-on	hands-off	4b–1	4b–1	Pl. N–Adv	Pl. N–Adv	=

handie-talkie	walkie-talkie	1–1	1–1	A–V	V–V	≠
hellseeking	job-seeking	4b–1	1–1	N–N	N–N	≠
hit list	hit man	3–1	3–1	N–N	N–N	=
hit squad	hit man	3–1	3–1	N–N	N–N	=
hophead	dopehead	3–4a	3–4a	N–N	N–N	=
hot war	cold war	4b–1	4b–1	A–N	A–N	=
hot warrior	cold-warrior	4b–1	4b–1	A–N	A–N	=
inner space	outer space	1–3	1–3	A–N	A–N	=
jungle fever	yellow fever	4a–4b	4a–4b	N–N	A–N	≠
kidflick	kidvid	3–3	3–3	N–N	N–N	=
kiteboard	surfboard	2–1	1–1	N–N	N–N	≠
last-second	last-minute	1–4a	1–4a	A–N	A–N	=
little-endian	big-endian	4b–6	4b–6	A–N	A–N	=
love-in	sit-in	1–5	1–5	V–Prep	V–Prep	=
low-rise	high-rise	1–5	1–6	A–N	A–N	=
metalhead	breadhead	3–4a	4b–4a	N–N	N–N	≠
meth head	dopehead	3–4a	3–4a	N–N	N–N	=
moonquake	earthquake	1–1	1–1	N–N	N–N	=
motormouth	big mouth	4b–4a	4b–4a	N–N	A–N	≠
mouse potato	couch potato	4a–4b	4a–4b	N–N	N–N	≠
mouse race	rat race	4a–4b	4a–4b	N–N	N–N	=
narrowcast	broadcast	1–3	1–3	N–N	N–N	=
orangefield	greenfield	4b–1	4b–1	A–N	A–N	=
petrolhead	breadhead	4a–4a	4b–4a	N–N	N–N	≠
Pink Friday	Black Friday	4b–1	4b–1	A–N	A–N	≠
pink-collar	white-collar	4b–4a	4b–4a	A–N	A–N	=
podcatching	podcasting	1–1	1–1	N–N	N–N	=
pothead	dopehead	3–4a	3–4a	N–N	N–N	=
potty mouth	foulmouth	4b–4a	4b–4a	N–N	A–N	≠
prooflisten	proofread	1–1	1–1	N–V	N–V	=
puppy leave	maternity leave	1–1	1–1	N–N	N–N	≠
read-in	sit-in	1–5	1–5	V–Prep	V–Prep	=
sandboard	snowboard	1–1	1–1	N–N	N–N	=
second-hand drinking	second-hand smoking	1–3–1	1–3–1	A–N–N	A–N–N	=
sleep-in	sit-in	1–5	1–5	V–Prep	V–Prep	=

slow food	fast food	3–1	3–1	A–N	A–N	=
small data	big data	1–1	1–1	N–N	N–N	=
smartglasses	smartphone	4b–1	4b–1	A–N	A–N	=
smartwatch	smartphone	4b–1	4b–1	A–N	A–N	=
smoke-easy	speakeasy	1–1	4a–1	V–Adv	V–Adv	≠
soft power	hard power	4b–1	4b–1	A–N	A–N	=
stomach share	market share	4a–3	1–3	N–N	N–N	≠
subway desert	food desert	1–4b	1–4b	N–N	N–N	=
tablet potato	couch potato	3–4b	4a–4b	N–N	N–N	≠
third-hand smoke	second-hand smoke	1–3–1	1–3–1	A–N–N	A–N–N	=
try-hard	die-hard	4a–1	4a–1	V–Adv	V–Adv	=
walkie-lookie	walkie-talkie	1–1	1–1	V–V	V–V	=
walk-in	drive-in	1–1	1–1	V–Prep	V–Prep	=
wallet share	market share	4a–3	1–3	N–N	N–N	≠
warm-down	warm-up	1–1	1–1	V–Prep	V–Prep	=
web rage	road rage	4a–1	4a–1	N–N	N–N	=
weddingmoon	honeymoon	1–4b	4b–4b	N–N	N–N	≠
white market	black market	4b–1	4b–1	A–N	A–N	=
white money	black money	4b–2	4b–2	A–N	A–N	=
work rage	road rage	4a–1	4a–1	N–N	N–N	=
work-hard	die-hard	4a–1	4a–1	V–Adv	V–Adv	=