Some new insights into the semantics of English N+N compounds

by

Elizaveta Tarasova

A thesis submitted to Victoria University of
Wellington in fulfilment of the requirements for
the degree of Doctor of Philosophy

Victoria University of Wellington
2013

Abstract

Some new insights into the semantics of English N+N compounds

This thesis focuses on English N+N compounds and the primary purpose of the study is to investigate the way in which compounded structures acquire their meaning and to check the way in which the semantics of each of the constituents contributes to the overall meaning of the structure. The way in which such contributions are made should be inferable from the linguistic analysis of the structure and meaning of compounds. In order to do this, the thesis looks first at the morphological productivity of the constituents comprising a compound. The second aim is to identify whether the productivity of a compound constituent on the morphological level coincides with the productivity of the semantic relation realised in the constituent family. The discussion of the results obtained from a corpus study provides plausible explanations for the regularities noted in the course of the analysis by using some of the relevant principles from the complex of approaches including the Construction Grammar and Cognitive Grammar approaches.

Examples of compounds were collected from the printed media (NZ broadsheets) and the BNC. The analysis of the data used both quantitative and qualitative methods.

The quantitative analysis of the data confirms two hypotheses: (1) that a constituent is more productive in just one of the positions (modifier or head)[1], and (2) the more productive a constituent is, the more likely it is to realise a single semantic relation in a constituent family. The qualitative analysis involves consideration of the semantic content[2] of the concepts in each constituent in order to see how this content is reflected in the semantic relations realised by a constituent. It is discovered that the semantic content of the head is a stronger predictor of the relation realised in a compound than that of the modifier.

The study is important in order to better understand the factors that govern the formation of compounds and the patterns that speakers use in the process of coining complex lexical items.

[1] The notion of headedness comes from American structuralism (Bloomfield 1933), where it is used for the distinction between endocentric constructions consisting of a head and a modifier, and exocentric constructions, which do not contain a head, e.g. *jailbird*, *hunchback*. In endocentric constructions, the compound is the hyponym of the head element, e.g. *apple tree* is a kind of tree; therefore, *tree* is the head of the construction. The element *apple* is a modifier. In English endocentric N_1+N_2 constructions, the modifier typically takes the position of N_1, and the head takes the position of N_2.

[2] In this research the term 'semantic content' is used to talk about the word's meaning as listed in dictionaries.

Acknowledgements

This doctoral thesis would not have been possible without the help and support of the kind people around me, to whom I owe my deepest gratitude.

Above all, I am grateful to New Zealand Government and particularly Education New Zealand Trust, for the award of a New Zealand International Doctoral Research Scholarship that provided the necessary funding for this project.

I am eternally grateful to my primary supervisor Prof. Laurie Bauer for his continuous support, for his patience, encouragement, critical advice, stimulating suggestions and unsurpassed knowledge of language matters. It is an honour for me to have had the experience of working under his guidance and I could not have imagined having a better advisor for my PhD study. I am also deeply indebted to my secondary supervisor Dr Paul Warren, whose objective comments on my work made me think about things from different angles, and who always made himself available when I needed his help and advice. I want to say special thanks to Prof. Ingo Plag, whose genuine interest in the results of my research and enthusiasm helped me regain confidence in myself in the final stages of this project.

I am obliged to the staff of the School of Linguistics and Applied Language Studies for all their help, support, interest, and promotion of a stimulating and welcoming academic and social environment. The kindness and optimism of my colleagues in the School and sharing their experience with me was very motivating and helped me a lot. I am also grateful to the School for providing me with the opportunities for professional growth and development by involving me in exciting research projects and teaching.

I would also like to acknowledge the financial, academic and technical support of Victoria University of Wellington and its staff. The library facilities and computer facilities of the University have been crucial for the completion of this project.

I would like to express my perpetual gratitude to my parents and relatives. My mum has always encouraged me, although letting me leave to do a PhD was probably the most difficult decision for her. And the example of my deceased father has been driving me throughout all these years. I am sure that he would have probably said that my research is totally impractical but would have been very proud of me anyway.

I greatly appreciate my fellow PhD students Kiet Ho, Natalia Beliaeva, Anna Piasecki, Brian King, Sharon Marsden, Elena Maydell, Yulia Khrypko, Peter Cooper and many others, who provided

valuable feedback at the meetings of our thesis group and whose friendliness made the study process less solitary.

Thank you to all the people that have become a very big part of my life and were there for me whenever I needed. Amongst all I would like to thank Nancy Ford and Nicole Randall, for being so kind and generous to me all these years and for putting up with me when things were especially tough. My special thanks go to my boyfriend, Gerry Keating, for being so patient and understanding.

I am also grateful to my friends and colleagues in Russia, whose continuous support was inspiring. I would like to say special thanks to A.A. Zalevskaya, my mentor and supervisor of the Candidate of Philological Sciences thesis, to whose academic and personal integrity I aspire. Thank you to Aleksei Gouryashkin and his mother Alevtina Gouryashkina for encouraging me to apply to do a PhD abroad and for all the help at the initial stages of this project.

I also wish to thank my examiners Dr Frank Boers, Prof. Koenraad Kuiper and Prof. Jim Miller for their insightful comments and constructive feedback on my work.

Thank you to Wendy Collings, who looked closely at the final version of the thesis for English style and grammar, correcting both and offering suggestions for improvement. For any errors or inadequacies that may remain in this work, of course, the responsibility is entirely my own.

Contents

List of Tables

List of Figures

Abbreviation key

Adj + N – adjective + noun

BNC – British National Corpus

COCA – Corpus of Contemporary American English

CWF – constituent word family

Gen. – Genitive case

N+N – noun + noun

NDVCs – non-affixal deverbal compounds

OED – Oxford English Dictionary

Pl. - plural

PR – probability rate

RDP – recoverably deletable predicate

Sing. – singular

1. Introduction

The formation of meaningful sequences consisting of two adjacent nouns (henceforth – N+N), e.g. *door handle, pencil sharpener, biology research, team management, singer-poet,* etc. in English (as well as in other languages) has been and continues to be the centre of attention of many researchers in a number of linguistic fields including morphology, semantics, cognitive grammar, computational linguistics, psycholinguistics, applied linguistics, linguistic typology, etc. Researchers within these different fields concentrate on the questions that are felt to be important for their particular goals in research in this area. As a result, a diverse body of literature concerning the structural peculiarities, typology, semantic characteristics, and processing patterns of N+N sequences is currently available. In the last twenty years a large amount of work has been done in order to understand how individuals construe the meaning of a compounded structure when they encounter them in the process of communication (mainly written), e.g. Estes (2003), Estes and Jones (2006), Gagné (2000, 2001), Gagné and Spalding (2004, 2006a, 2009, 2010), Wisniewski (1997), etc. However, not many researchers have looked into the issues associated with the process of forming N+N sequences. This oversight applies to the empirical testing, actual modelling of the use of these structures and the descriptive notion. Work that examines the production of nominal compounds has been quite patchy and disjointed (Clark, Hecht and Mulford 1986, Downing 1977, Levi 1978, Windsor 1993) and the explanations that the existing theoretical frameworks offer do not provide a solid base for further studies. This may be because the production literature to date tends to be mainly concerned with issues other than creation/coining of these units *per se*. For example, Clark et al. (1986) and Windsor (1993) were primarily concerned with production of N+Ns as an aspect of language development, while Downing (1977) and Levi (1978) focused on debating taxonomies of relations within N+N sequences. It is also very hard to generalise over the existing studies because of their widely differing methodologies, from speech error analysis to picture naming and from context-free production to linguistic analysis.

This thesis discusses research concerned with English endocentric structures consisting of two adjacent nouns[3] and especially concentrates on the matters associated with their semantics. It has long been observed that N+Ns in English as linguistic forms combine two or more linguistic parts into a semantic whole, though they do so without there being any grammatical marks to indicate

[3] In my research I follow the most commonly accepted definition offered by Bloomfield (1933), according to which endocentric compounds are hyponyms of their head elements. For example, *apple tree* is a kind of tree, *cookie jar* is a kind of jar, and *sun glasses* are a kind of glasses).

what the relation between these two (or more) parts is and how this relation has been obtained. The position that their meaning is listed in the lexicon and is accessed as needed simply cannot hold true in all cases due to the high productivity of N+N units as a morphological process. The ability of a reader/hearer to unpack the semantics encoded in compounds and that of a writer/speaker to use this information for coining of novel compounded items is intriguing because the surface form of such compounded unit does not reliably disambiguate its meaning. If we combine *bear* and *paw*, we get *bear paw*, and we understand that the relation between the components is that of possession, i.e. a bear HAS paws. If we change the second constituent, e.g. into *scare*, the relation between the constituents changes and the bear is understood as the cause of the scare. The same happens if we change the first constituent, e.g. *citizen army* vs. *liberation army*; *security police* vs. *city police*, *jewellery box* vs. *cardboard box*. However, formally, we are still dealing with the combination of two nouns and there are no markers to indicate the changes in meaning. Moreover, the fact that one and the same noun may occur with different relations suggests that we cannot make a reliable prediction as to what relational meaning will be realised with a certain noun constituent. Ó Séaghdha (2008) points out that in order to draw conclusions about the regularities for nouns to be used in N+N sequences it is necessary to approximate the kinds of conceptual knowledge that humans possess and how this knowledge reveals itself in linguistic structures. So, we should be looking at approximations and probabilities rather than strictly observed rules.

As mentioned above, sequences that consist of two nouns are the objects of this research and I do not look at the other nominal constructions like Adj+N (*greenhouse*, *blackbird*), Adj_{Assoc} (*atomic bomb*, *medical student*), N's+N (*mother's smile*, *dog's breakfast*), etc. Nominal constructions like these possess a number of characteristics that make them in some respects similar to N+N sequences. For example, Bauer and Tarasova (to appear) demonstrate that the same set of semantic relations recurs more or less consistently with different constructional patterns. Based on their analysis, in all of the following cases the instrumental relation unites the constituents and all of the sequences can be interpreted using the N2 USE N1 pattern: *wind farm, manual labour, car's driver, driver of the car, electrocardiograph, jazzercise, bus and train services*. Although the research that would look into the similarities of the semantics in different nominal sequences is by all means interesting, I feel that each of the constructions deserves a separate investigation.

Another reason for limiting the scope of the analysis is connected with the unmarkedness of the first element. The qualitative characteristics of the attribution performed by the first noun in an

N+N endocentric sequence cannot be calculated based on the morphology of the constituents. The only information that we gain from the constructional features is that the first element modifies the second in some way; however, the quality of this modification is unpredictable from the structural characteristics. It is, therefore, interesting to know what factors are important for the generic grammatical relation to acquire a rather concrete sense in a given N+N sequence. So, this thesis addresses the issue of semantic relations that hold between the constituents of N+N constructions in detail.

The second specific feature of N+Ns is that they have a special meaning of their own, i.e. the meaning of a sequence is not equal to and is even sometimes relatively independent of the meaning of its constituents. This is typical of well-established items (i.e. lexicalized compounds that have conventionalised meanings) and novel N+N coinages. For example, in the case of such well-known compound as *letter box*, we deal with not just any box that has letters in it (although it may well be any box used for this purposes), but likely a box of a special design is meant. If we take such a novel compound as *a mouse jumper*, chances are that even if we have not encountered it before, we can still have some idea of what the combination may refer to. Based on our knowledge of jumpers, mice and how they can be related in extra linguistic reality, we may assume that a *mouse jumper* is more likely to be 'a jumper that has a picture of a mouse on it' (as in *bear jumper*) or 'a jumper that a mouse (maybe a toy one) wears' (as in *dog jumper*), rather than 'a jumper that is made from mice (wool)' (as in *alpaca jumper*) or 'a jumper that a mouse knitted'[4]. As can be seen from these examples, the differences in the interpretations of one and the same sequence are usually believed to be connected with the different semantic relations that can be inferred between the constituents.

However, even if we assume that the semantic relations are responsible for the reading a sequence receives, it is not clear how and where from these semantic relations appear and where they originate. Can we suggest that one or the other constituent is responsible for that, and if yes, which constituent is it and in what way? In other words, does the MADE OF relation in *glass house* depend on the semantics of *glass* or *house*, on the fact that *glass* modifies *house*, or something else?

[4] Based on search in Google (22.03.2013) a *bear jumper* is most likely a jumper with a picture of a bear on it, a *dog jumper* is a jumper that dogs wear, an *alpaca jumper* is a jumper made from alpaca wool; and in this reality it is highly unlikely to see an animal knitting a jumper at all.

Another question is what makes one interpretation more acceptable (and, consequently, more accepted) than the other? Even in the cases with lexicalised[5] items (i.e. those that have canonised entries in lexicons) like *cardboard box*, at the moment of coinage there should have been something that ruled out the reading 'a box where one keeps cardboard' (as in *shoe box*) making the reading 'a box made from cardboard' (as in *plastic box*) conventionalised. The literature on the topic does not seem to provide a comprehensive answer to this question.

The present study makes an attempt to investigate these questions and suggest some possible explanations as to how N+N sequences acquire the meaning they have and what is the role of each of the constituents in meaning construal of the analysed structures. This means that this research aims at the analysis of the way compounds are produced. However, neither linguistics nor cognitive science currently possesses efficient procedures that would allow for making judgements as to what is going on in the human brain in online speech production. All the currently available methods that test online production of compounds (picture naming, producing compound words from definitions) are often ineffective and the interpretations of the results obtained from such studies often sound somewhat arbitrary. Therefore, I am not looking at online speech production but at the corpus data, which are viewed as the final stage or outcome of production processes.

In the course of my analysis I will also look at the findings obtained in psycholinguistic research on how N+N sequences are perceived by individuals and I will see if this data can be used for the explanation of the patterns noted in the formation of such structures. Undeniably, comprehension is an integral part of production without which the latter could not occur because a novel sequence will be produced on the assumption that it is comprehensible for the listener/reader. The model offered by Levelt, Roelofs and Meyer (1999) provides evidence demonstrating how speech comprehension and production are connected. For example, according to Levelt et al. (1999), speakers make use of their speech comprehension system to listen to and process their own speech in exactly the same way they listen to and process the speech of others.

Although processes involved in speaking and understanding language are fundamentally different, both recruit representations of conceptual-semantic, syntactic, morphological, and form information. As neuroscientists (e.g. Buchsbaum, Hickok and Humphries 2001, Hickok, Houde & Rong 2011, Wise, Chollet, Hadar, Friston, Hoffner and Frackowiak 1990) suggest, speech perception and speech production have a number of similar features. Firstly, both speech

[5] The issue of lexicalization is discussed in Section 2.5.1

perception and speech production involve the same neural systems, e.g. auditory, motor, control, etc. (Hickok et al. 2011). Secondly, single word comprehension is reported to engage the same (Wernicke's) area in the brain as single word comprehension at the stage of silent word generation (Wise et al. 1990). Thirdly, some studies demonstrate that there is overlap in neural systems that are believed to participate in speech perception and speech production (Buchsbaum et al. 2001).

Since, as pointed out, there are limitations in terms of the availability of tools for testing speech production processes, we can look at the final results of speech production, i.e. N+N coinages as they occur in the language. In order to make some valid speculations concerning the way these structures are coined, we can search for some regularities in paradigms that use certain nouns as constituents of compounds. Such paradigms can firstly be analysed from the viewpoint of the structure of the items comprising them. For example, one of the issues that seems to be overlooked in research in this field is the issue of the productivity of nouns as constituents of compounds. It is generally believed that compounding is one of the most productive word-formation processes in English, but there is hardly any research that looks at the productivity of individual members of compounded structures. In addition, it rapidly becomes clear from data that certain nouns have a tendency to realise only one aspect of their meaning when combined with other nouns. This may result in the consistent realization of one and the same relation in the paradigm, which also can have some influence on the choice of the second noun in the construction. For example, nouns denoting material (*stone*, *brick*, *concrete*) tend to be used with artefacts (*house*, *building*, *bridge*) that are normally made from these materials. Therefore, the analysis of the semantic relations within paradigms of compounds as well as the analysis of the semantics and pragmatics of compounds and their constituents are important for this research.

So, in order to investigate the questions of the influence of constituents of compounds on the meaning of the whole coinage I look at the regularities in the formation of compounds from the viewpoint of their structure and at the regularities of the semantics of the coinages. The analysis of both serves as the basis for the discussion of the mechanisms and factors of compound production.

In my research I use the data collected from the NZ printed media and the BNC; and I use dictionary definitions in the analysis of the semantics of N+N compounds and their constituents. Chapter 2 provides a review of literature relating to research in nominal compounding in English relevant for the current study and provides a discussion of the problematic issues in compounding research. Chapter 3 describes the methodology of the study of English N+N compounds and

provides the results of the global trends noted for the collected data. Chapters 4-5 look into the data in more detail, analysing the behaviour of individual constituents in compounds. In Chapter 6 I draw the conclusions and talk about the noted trends in the morphology and semantics of the analysed units. I end the thesis with a brief recapping of the findings and implications for further research in the field.

2. Literature review and theoretical prerequisites

2.1 Preamble

As pointed out in the *Introduction*, the current investigation looks into the issues underlying the formation of constructions consisting of two adjacent nouns, e.g. *client computer, conference plenary, community issues, money problems, linguistics student*. Constructions of this type seem to cause a number of problems for linguistic analyses in different frameworks. At the same time they do not seem to cause any problems to the users of English who actively form sequences of this type and operate with them extensively in everyday communication.

There are several questions that are traditionally associated with such constructions. The most debated one concerns the status of such units in the system of the English language. Sequences of this kind take an intermediate position between morphology and syntax since they demonstrate features that make it possible to view them as creations of either of the levels of English grammar. Units of this type are not always listed in dictionaries and can be constructed on the spot if needed. Native speakers do not seem to have difficulties with accessing their meaning, even if they have never encountered them before, so they are novel to these individuals. At the same time, it is hard to refer units of this type to the category of novel constructions, since some of them may receive relatively high counts in corpora of the English language and may have existed in the language for some time and, besides, they do not necessarily utilise metaphor or metonymy, the criterion often used for making judgements on compounds' creativity (see Benscez 2006a for details).

There have been several attempts to develop a set of rigid criteria that would allow linguists to unequivocally decide whether a given N+N sequence is a compound or a phrase and whether they should be analysed as morphological or as syntactic units. An extensive body of research in this area has been developed over the years, and this research has been recently enriched by the developments in psycholinguistics and cognitive science. Interestingly, the analysis of the psycholinguistic literature on the subject demonstrates that in the majority of works the distinction between N+N compounds and phrases is not taken to be important by the researchers. The questions that current studies attempt to answer mainly concern the matters of meaning construal. Therefore, experimental studies described in the literature aim at the analysis of online processing of meaning of N+N sequences, which are often called 'conceptual combinations' by psycholinguists and cognitive linguists. The modern approaches are usage-based and they attempt

to analyse language structures from the position of the user of the language. As a result, the focus of the linguistic analysis has shifted and the processes underlying sense creation of the linguistic expression by an individual seem to have gained priority.

Another reason for the lack of attention to the distinction between compounds and phrases in modern usage-based frameworks may be connected with the following. In order to eliminate ambiguity in the interpretation of the results and to reduce the number of factors (e.g. the degree of familiarity of an item to different speakers, personal attitudes to objects/phenomena named by an N+N sequence, etc.) that may influence interpretation of N+N sequences by the subjects of the experiments, researchers often use artificially created sequences that either do not exist in the language or have a very limited sphere of usage, e.g. *zebra horse, canary crayons, porcupine mushroom, bumblebee moth*, etc. The likelihood that such items would ever be coined in authentic language use is very small, but since psycholinguistic and cognitive linguistic research is concerned with other matters, this issue is disregarded and all N+N items are considered to be compounds as a default.

In this research I want to look at N+N units that occur in real. These units seem to be overlooked in the literature, but I find them interesting for the analysis, as their characteristics allow for looking at combining two nouns as a process of coining new units that may or may not become part of the lexicon. By looking at the formation of such items, we may be able to see how meaning is assigned to them and what factors may be of importance. I do not want to get into the discussion of whether these units are compounds or phrases, as I do not feel that the distinction between N+N compound nouns and phrases is essential for my analysis. However, since syntactic and morphological occurrences of N+N structures are often not distinguished terminologically in the literature and since most of the discussion of these items is carried on under the heading of "compounds", I will use the term "compound" for convenience of presentation. I will also look at the literature on nominal compounds in order to discuss the issues relevant for my research. Below, I also outline some of the questions that concern the structural peculiarities of N+N sequences in English, but, since this research is mostly concerned with the semantics of the analysed units, I mainly concentrate on the issue of semantic relations of N+Ns and their role in sense creation of compounded structures. In the course of the review of the literature on the topic, I provide a discussion of the strengths and weaknesses of some of the approaches relevant for this research, and conclude the chapter with an outline of the chosen framework.

2.2 Limiting the scope of the study

A number of studies are devoted to nominal compounding (especially with compounds made up of a sequence of two nouns), which is one of the most productive patterns of extending the vocabulary of modern English.

There are several classes of compounds in Modern English and a large number of works are devoted to the issues of their classification and their place in the English word-formation system. One of the most consistent (in my view) classifications is presented in Bisetto and Scalise (2009).

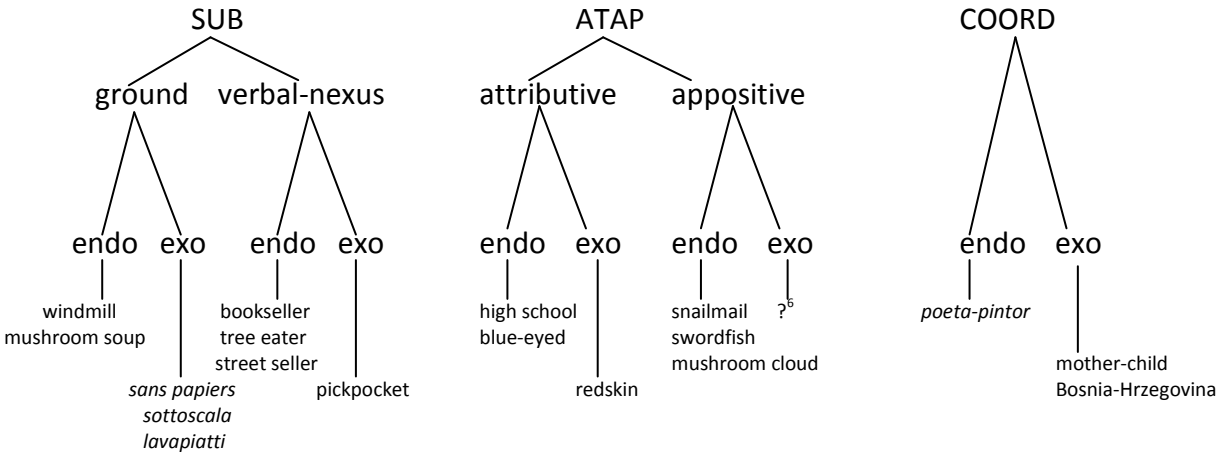

Figure 2.1 Bisetto and Scalise's (2009: 50) classification of compounds

Their classification has three grand levels of specification. At the macro-level all nominal constructions are subdivided into three big classes based on the grammatical relations between the constituents: subordination, attribution and coordination. Based on these fundamental relations, as shown in Bisetto and Scalise (2009), the differences in the more specific interpretations (i.e. semantic/interpretative relations) are at the second level of classification related to the categorical nature of compounds. Each of the second level classes is further divided into two subclasses based on whether they are endocentric or exocentric. In endocentric compounds one of the elements (typically the right-hand element in languages like English) is the head of the sequence. Bloomfield (1933) identifies headedness through hyponymy, i.e. the compound as a whole is the hyponym of its head. As Bauer (1990) shows, it seems that hyponymy is the least controversial criterion; hence, I will use hyponymy as a determining criterion for headedness. Typically, in English endocentric compounds are hyponyms of their right-hand element (Bloomfield 1933: 235). For example, *apple tree* is a kind of tree, *cookie jar* is a kind of jar, and *sunglasses* are a kind of glasses. Exocentric compounds, are not hyponyms of any of their

[6] Bisetto and Scalise (2009) note that appositive exocentric compounds are not well represented in their database.

constituting elements, and thus seem either not to have a head or their head is external to the compound itself; for example, *redhead* refers to a person with red hair; similarly, *scarecrow* is a thing that scares crows away, and *muffin top* refers to a roll of fat visible above the top of a pair of women's tight-fitting low-waisted trousers. The meanings of such constructions are based on metaphoric or metonymic uses of the head noun (Benczes 2009: 49). Therefore, we cannot say that exocentric compounds are hyponyms of their heads because the head noun is used in a figurative meaning.[7] However, as noted by Fabb (1998: 67) one and the same compound can be considered to be endocentric and exocentric at the same time depending on the interpretation and the opinion of the analyst. Fabb (1998: 67) gives the example of *greenhouse* whose being endocentric or exocentric depends on whether you think it is a kind of house or not. Similarly, quite often it is difficult to say whether a given compound is attributive or subordinate as the degree of relativity/attributivity may vary depending on a number of factors, including context, conventionality of usage, etymology, homonymy etc. In theory, *a cricket bat* could be a flying mammal that eats insects that sing, rather than a club for playing a sport with. Another example is *lemon fish*, which fits the category of subordinate compounds when interpreted as 'fish cooked with lemon'. However, understanding *lemon fish* (cobia), as the name of the species (Rachycentron canadum) yields a different categorization. At first sight it appears to be an attributive compound, but cobias do not seem to share any similarities with lemons.

So, it seems that there are cases that are difficult to refer to one and only one category, which means that the boundaries between the categories of this classification are fuzzy. In my research I am interested in subordinate endocentric compounds, which Bisetto and Scalise (2009) call 'ground', and which are also termed 'root' compounds. As the current research focuses only on one type of compounded structures, I will not go into much detail and discuss the differences between different classes and subclasses here, nor will I attempt to analyse the classification in much detail. One and the same compound can receive different readings depending on a number of factors, and this fact should be considered in the linguistic analysis. However, even in novel cases, some meanings seem to be more available than others. For example, a compound *?mouse murderer* is more likely to receive an agentive interpretation of 'a person who murders mice' rather than 'a person who uses mice as a tool for murdering' (c.a. *axe murderer*). At the same time, *?machete murderer* is likely to use a machete for murdering. There are novel cases where the interpretation may vary depending on the context. For example, the compound *?tiger*

[7] It is in dispute though whether metonymic interpretation is a good reason for relegating items like *redskin* and *greybeard* to grammatically headless constructions (see Carstairs-McCarthy (2010) for discussion).

murderer, based on its written form, can be understood as 'a murderer of tigers' as well as 'a tiger that is also a murderer'. I believe that irrespective of the degree of availability of one interpretation over the others, both understanding and creating such compounds involves active processing of the meaning of the constituents as well the meaning of the whole, as well as the analysis of the discourse context. Highly lexicalised items are generally believed to not require deconstruction and are accessed (and produced) as simplex words (see, for example, Bertram and Hyönä 2003, Frisson, Niswander-Klement and Pollatsek 2008, Pollatsek and Hyönä 2006, Pollatsek, Hyönä and Bertram 2000, Schreuder and Baayen 1997, van Jaarsveld and Rattink 1988). Another difficulty with classifying compounds concerns deciding what should be considered a compound and thus be included into the classification, and what should be left out. For example, one of the criticisms of this classification is that it does not seem to provide a niche for blend-compounding, for example, *brunch* (*breakfast* + *lunch*) and *smog* (*smoke* + *fog*). This brings us to the question of what we can call a compound, which is discussed in the next section.

2.3 Difficulties in distinguishing compounds from other strings

As mentioned above, one of the most widely discussed issues in compounding research for the last two centuries is how to decide whether a sequence of two nouns is a lexical unit (compound) or a syntactic one (phrase). A large number of works are devoted to this question and I will provide a very brief outline of the main approaches to it.

An analysis of literature on the topic shows that the problem of defining a compound is approached differently by different authors. Practically all the definitions of compounds can be divided into two types: grammar-oriented and semantics-oriented (although, of course, this is an artificial dichotomy, as form and meaning are two sides of the same coin). One of the examples of the first type of definitions is that of Jespersen (1942: 134), who says that it is insufficient to base a definition entirely on meaning in a morphological treatment of compounds, and, concentrating on the functional side, believes that a compound may be defined "… as a combination of two or more words so as to function as one word, as a unit". Marchand (1960: 11) focuses on the grammar of the constituents and talks about the necessity of two independent lexemes to be combined into a morphological unit for that unit to be considered a compound. This definition implies that once combined into a compound; the constituents stop being independent lexemes and "go one level down" in the language hierarchy. Using the terminology of Halliday's (1976: 70) systematic approach to grammar, we can say that this is an example of rankshift, which, when applied to the

category of class, runs in one direction from highest to lowest unit. This means that the unit of language of any rank can be used as a unit of lower level.

> … it may happen that, while the "normal" unmarked rank-value[8] of the item operating as a constituent at a given place in structure is rank x, we also find instances where an item of rank *y* occurs instead… The formulation "*x* used as *y*" implies that the terms appearing in it are useful precisely because they are not merely labels for morphological types but have certain characteristic functions associated with them. In cases of (downward) rankshift, an item normally having the function of (entering into the paradigmatic and syntagmatic relations associated with) rank *x* characteristically "loses" these functions on taking over those of rank *y*… (Halliday 2002: 122).

So, according to Halliday, the constituents of compounds no longer act as words (lexemes), but acquire the function typical of a lower rank unit, which is obviously a morpheme. At the same time, Halliday's (1976) theory permits 'shunting' (moving up and down the rank scale), thus permitting the full description showing the interrelation of the categories. This means that even being rankshifted in the downward direction, components that constitute a compound can be analysed as words even when they function in a morpheme-like manner. This sounds very much like Alison Wray's view of formulaic sequences (2002), later re-labelled 'morpheme-equivalent units' (Wray 2008), to which compounds clearly belong in her framework. According to her theory, people will process them holistically but can also break them up into parts when the need arises.

Another commonly used grammar-oriented definition is offered by Bauer (1978: 54), who combines the grammatical and functional sides of a compounded unit and, at the same time, coincides with the idea of rank shifting: "… a compound is a morphologically complex unit, made up of two or more words (lexemes) acting as a single word (lexeme)". The problem with this definition is that it does not make it clear whether "acting as a single word" makes a morphological unit a word or do we still deal with two lexemes that happen to occur together and that function in the word-like manner. Another issue with this definition is that it remains unclear whether "acting as a single word" also implies being semantically similar to a single word, i.e. have a prescribed meaning that is listed in the lexicon.

Bisetto and Scalise (2005) define compounds in the following way: [X $\mathfrak{R}$ Y] Z, where X, Y and Z are lexical categories and $\mathfrak{R}$ is the implicit relation between the two constituents (a relationship that is not explicated by any lexical item). The nature of $\mathfrak{R}$ is presumed by the authors (who seem to

[8] The notion of "rank" is one of the central notions in Halliday's grammar and it is used to refer to a position of a linguistic unit on an abstract scale (rank scale) which represents the taxonomy of the language structure. In Berry (1975: 104) an example of the simplest rank scale is given, where a sentence gets the highest rank on the rank scale and a morpheme gets the lowest rank.

follow Bloomfield 1933 and Marchand 1969) to be grammatical because of the differences in the interpretational patterns of such compounds as *notebook* and *doctor-poet* ('book **for** notes' and 'doctor **and** painter' respectively). However, it is clear that it is impossible to completely divorce grammar from semantics, since (as Construction Grammar claims) all constructions (small to large) are symbolic units and thus carry meaning.[9]

This definition is aimed at covering compounds in different languages and disregards the morphological status of the base elements comprising a compound. As a result, it is possible to account for a variety of constructions including those in which the constituents are not orthographically represented as independent lexemes, e.g. *cranberry*, *cobweb*, *twilight, astrophysics, horticulture*.[10] The main problem with this definition, however, is that it implies that all compounds are binary, i.e., consist of just two nouns but, under this assumption, constructions like German *Rot-Schwarz-Gold* 'red-black-gold' or French *bleu-blanc-rouge* 'blue-white-red' either cannot be referred to compounds or there is a problem with their internal bracketing. Another issue is that, postulating the nature of $\mathfrak{R}$ as grammatical gives grounds for some questions. On the one hand, it is clear that the differences in interpretations of combinations like *singer-doctor* vs. *tulip tree* and *pot plant* arise from the differences in their grammar, with the last two examples being endocentric constructions and the first one not (or, at least, not necessarily). But, if we take combinations within one classificatory category, it is not clear how grammar can account for the differences in the meaning of combinations like *apple tree* and *apple juice*? As it turns out, we can define a compound grammatically but 'grammar'-based approaches may only help define compounds as different from other combinations. At the same time, they fall short of explaining the precise semantics of the compounds, i.e., why a given compound means what it means. This is

[9] Other approaches suggest a semantic nature for $\mathfrak{R}$ based on the processes of concept-combination. Assuming the semantic nature of $\mathfrak{R}$, it is possible to say that it is determined by the cumulation of the semantico-encyclopaedic information associated with the two constituents. According to Pustejovsky (1995), the interpretation of endocentric N+N compounds depends on the possibility of associating the different semantic pieces of information in the *qualia structure* of the two constituents. The different interpretations allowed by these compounds depend on the possibility of varying the association between the features of the respective *qualia structures*: these features are elicited from the context in compounds can be found. Scalise and Bisetto (2009) reconsider their claim about the grammatical nature of the relation and say that $\mathfrak{R}$ is a semantic relation between the two constituents determined by the semantico-encyclopaedic information associated with the component lexemes. They also point out that it is the information of the head constituent that selects a piece of information of the non-head, thus determining the interpretation of the compound.

[10] As claimed by the authors, this definition is aimed to cover a large variety of compounds found in different languages, especially those in which compounding mainly employs bound root morphemes.

the question that the current thesis aims to answer. Compounds are understood here as conceptual units[11] rather than merely a combination of two or more words/roots/stems/lexemes.

So, at one end of the spectrum, we have grammar-based definitions, which do not tell much about the meaning. However, defining an N+N compound from the viewpoint of the peculiarities of its semantics seems to represent another extreme. Although the semantics of compounds is definitely intricate, a definition based on semantics alone can run the risk of confusing compounds with other language units (e.g. idioms). An example of this can be seen in Kruisinga and Erades' (1953: 385) work, who claim that "[T]wo or more words are said to be fused into a compound when the meanings of the component parts together no longer indicate the meaning of the whole word". Based on this definition, it seems that we should put compounds into the category of idioms because the meaning of the constituents cannot be traced from the meaning of the whole structure. It seems that this definition can be applied to a limited number of examples that are not hyponyms of their head elements (hence are not examples of endocentric compounds) like *redhead, pickpocket, jailbird, fleabag, phoneneck, muffin top*, and are formed on the basis of metaphor or metonymy. Based on the classification provided in Section 2.1 (Bisetto and Scalise 2009: 50), such compounds should be classed as exocentric and can be conventionalised and/or lexicalised to such a degree that their meaning cannot be deduced from the meanings of the elements, but has to be learnt for the unit as a whole. However, such compounds are exocentric, which are not very common in English. Instead, we normally deal with coinages that are endocentric[12] and relatively transparent, i.e. it is possible to see traces of meaning of each component in the meaning of the whole coinage. If we encounter a novel compound that we had never come across before, these traces of meaning allow us to make a sensible guess as to what the meaning of this item may be. As suggested by Relevance Theory (Sperber and Wilson 1986), such guesses cannot be based on the semantics of the constituents alone but must be supported by contextual cues, whatever the nature of these cues is. A number of compounds that are lexicalised and utilise metaphor or metonymy can still be characterised by a relatively high degree

[11] The issue of defining what a concept is in linguistics has become a popular topic of investigation recently, especially with the development of the Cognitive Grammar approach. A number of authors have looked into this issue, e.g. Clark (1993), Croft (2004), Dillon (2000), Evans (2009), Likhachev (1997), Stepanov (2007) Talmy (2000), Wierzbicka (1980), to name just a few. In this research I assume Stepanov's (2007) position that a concept is a cognitive unit of meaning, an abstract idea or a mental symbol, and can be defined as a "unit of knowledge", built from other units which act as a concept's characteristic .

[12] Scalise and Vogel (2010: 8-9) show that right-headed compounds are by far the most prevalent type of compounds; especially in Germanic languages, with around 87% of occurrences out of all compound types.

of semantic transparency[13], e.g. *gaslight* (the head constituent is metonymy-based, since it stands for a lamp that produces light, i.e. RESULT – CAUSER metonymy). So, it seems that in the case of semantic transparency we deal with a scale that goes from semantically opaque to at least partly semantically transparent items. Overall, the question of semantic transparency is one of the crucial issues in compounding research. I will, therefore, refrain from going deep in detail on this matter here, but will go over the aspects of this issue relevant for my research in Section 2.5.3.

Even in the case of semantically non-transparent or occasional items both of the constituents contribute some aspects of their meaning to the whole. Benczes (2006a, 2006b, 2009), for example, demonstrates how the meaning of the constituents can be traced even in the case of what she calls "creative compounds", i.e. those that utilise metaphor and metonymy. One of the examples she provides (Benczes 2006b) is the compound *knee-mail*[14] 'prayer'. In this case the modifying element *knee* describes the whole scenario, in which the instrument *knee* (kneeling while praying) stands for the action of praying and the head element *mail* is used on analogy to email, with God being understood as the receiver. A more extreme case is illustrated by the compound *jailbird* 'person serving a prison sentence' (Benczes 2006a). Here the head noun realizes a metaphorical meaning, i.e. an imprisoned person is associated with a caged bird. The metaphor 'A PRISONER IS A CAGED BIRD' allows for further associations between the modifier *jail* and *a birdcage*.

In more transparent cases the meanings of the nouns used as constituents of compounds are still somewhat different from the meanings of these nouns taken separately because different aspects of their semantics are accentuated in different combinations. For example, the noun *milk* used in the modifier position in *milk tooth* is different from *milk* in *milk bottle*, however, both of them contain traces of meaning we can find in MILK taken as an isolated concept. In the same way the noun *mill* used in the head position in *windmill* is different from *mill* in *pepper mill*, but still both bear the features that allow us to define them as kinds of MILL.[15] This partial desemanticisation has the effect of leading to the compound as a whole not having in actual usage the full set of potential meanings that might be attributed to the combination of elements in isolation. This is often seen in the literature as implying that the combination is formally and logically equivalent to

[13] In the most general understanding in semantically transparent items the meaning of the entire string can be derived from the combination of the meanings of its constituents. See Section 2.4.2 for discussion.

[14] This example is somewhat ambiguous and can be considered a type of a linguistic blend rather than an N+N compound. However, Benzsces (2006b) analyses it as a nominal compound.

[15] However, we cannot consider this to be a feature idiosyncratic to compounds; the same kind of change can be observed in the meaning of unmistakable phrases, e.g. *white* in *white coffee* is different from *white* in *white cloud*.

a word and functions in a word-like manner, something which is often referred to as 'the unity of meaning' or 'semantic unity' (Sweet 1900, Warren 1978).

It becomes clear from the literature that there is a commonly accepted opinion among linguists that N+N compounding in English does not fit into standard categories either of syntax or morphology. As noted by Libben (2006), this is because compounds are structures located at the crossroads between morphology and syntax, as they reflect both the properties of linguistic representation in the mind and grammatical processing. The argument for the existence of a putative distinction between N+N sequences that are compound words and N+N sequences that are phrases has continued in different spheres of linguistic research for a rather long time without any unambiguously definitive results.

There have been a number of attempts to develop some formal criteria that could help to divide all N+N sequences into compounds and phrases. Some of the criteria are as follows:

- compound stress pattern (Marchand 1960);
- the unity of semantic content (Sweet 1900, Bloomfield 1933; Smirnitsky 1956);
- the character of components' combinability (Koziol 1937);
- the order of the components (Bloomfield 1933);
- continuity (Marchand 1960; Bloomfield 1933), etc.

The formal criteria traditionally used for the analysis of N+N structures in order to differentiate them from compound words are called in question by a number of researchers: Arnold 1973, Bauer 1983, 1998, Bloomfield 1933, Levi 1978, Plag 2003, Smirnitsky 1956, Warren 1978, to name just a few. Additional criteria have been offered (the possibility of the first component replacement, idiomaticity of meaning, recurrence in speech, etc.) but all of them act selectively and cannot be applied to all attributive sequences of this kind. Besides, for every new criterion offered there are a number of counterexamples that easily overthrow carefully constructed theories. Bauer (1998: 78) shows how application of a number of tests for compoundhood traditionally used in research may yield controversial results for one and the same compound. So, many western scientists simply adopt the criterion of orthographic separation in identifying N+N sequences (e.g. Biber, Finegan, Conrad et al. 2000). Although for the purposes of Biber et al.'s (2000) research the choice of this criterion was determined by operational needs arising in the course of the analysis of huge corpora, it is necessary to point out that orthography of N+N units is not strictly set in modern English. Talking about spelling being used as a criterion Szymanek (1988: 41) sums up the argument saying that "... spelling convention for compounds cannot be taken

seriously…the orthography of English compounds is notoriously inconsistent: some compounds are written as single words (*postcard, football*), in others the constituents are hyphenated (*sound-wave, tennis-ball*), and in still others the constituent elements are spaced off, i.e. written as two separate words (*blood bank, game ball*)." I can add that we can also notice how one and the same compound can be listed in different orthographic form by different dictionaries. For example, Collins (Collins online) and Longman (Longman online) list both one-word and two-word spellings of the compound *coffeepot*, whereas Oxford (Oxford online) provides only the two-word spelling and Webster (Webster online) only one-word spelling. In the case of the compound *rainforest*, Oxford (Oxford online) and Collins (Collins online) provide only one-word spelling, Webster (Webster online) only lists it in two-word spelling, and Longman (Longman online) gives two-word spelling and hyphenated spelling. So, there is no consistency either in the way different dictionaries present one and the same compound, or in the way each single dictionary deals with the issue of spelling of different compounds.

Since the current research is not aimed at solving the issue of distinguishing between a compound word and a phrase, but is more concerned with the process of combining two noun concepts into a single meaningful unit, I will not discuss the advantages and drawbacks of each of the criteria here. In this research I accept the opinion shared by some linguists (e.g. Bauer 1978, 1983, 1998, Olsen 2000, Bell 2005, 2011, Bell and Plag 2012, Warren 1978) that there is not much evidence for a class of syntactic N+N constructions. Following the position outlined in Bell and Plag (2012), I use the term 'N+N compound' for all N+N constructions consisting of two adjacent nouns, in which one modifies the meaning of the other, or where together they have a single meaning different from the meaning of either constituent individually. However, these exclude proper names (e.g. *Laurie Bauer*), and constructions with an appositive modifier (e.g. *lady Jane*).

2.4 Introductory comments on the semantics of N+N constructions

One of the specific features of N+N compounds that has been most commented on by linguists over the years is their semantics, namely the range of semantic relations expressed between the constituents, in the absence of any grammatical markers that would indicate in which way these relations is to be understood. For example, in the case of the *sleeping pill* sleeping is encouraged/induced by the pill; however, in *seasickness pill* or *antihistamine pill*, neither seasickness nor antihistamines are the desired effects (examples from Bauer (1979: 45)). In the

same way, *fire house, frame house, glasshouse, henhouse, town house* all show different relations between the first element of the compound and the *house* element.

The lack of overt grammatical markers indicating the particular reading a given compound will receive leaves room for a large number of different interpretations of one and the same unit (if one does not yet know the intended, conventional meaning). For example, *police car* can be understood as 'a car used in the police' or as 'a car designed for the police', but in practice this ambiguity is either avoided or does not seem to present any problems for communicants. Moreover, the ease with which new N+N compounds are formed in everyday usage as opposed to the ease with which new derivates (e.g. *beautiful*, *management*, *unhappy*) are formed (since the formation of compounds is not as constrained by linguistic factors) suggests that the relations between elements of a compound are more transparent than the relations existing between the elements of derivatives. Looking at children acquiring their native language, Tuggy (2005) suggests that they demonstrate a strong tendency to replace adult derivatives by more transparent items (e.g. a *dishwashing machine* instead of a *dishwasher*, *open-man* or *open-thing* instead of *opener*). Sweet (1900: 449) notes on this matter that:

> ... it is only by leaving open the logical relations between the elements of compounds that we are able to form them as we want them without stopping to analyze exactly the logical or grammatical relations between the words we join together, as we might have to do if we connected them together by more definite means, such as prepositions or inflections.

The presence of semantic relations between the constituents of a compound is sometimes used as one of the criteria for distinguishing true compounds from phrases. The meaning link between elements that form a compound must be intimate, irreversible, and permanent or at least habitual rather than the momentary, one-time-only, relation implied between the elements of a phrase. For example, *field mouse* can only refer to mice that are strongly associated with a field, such as living in or around it, and not just casual, situational or accidental connection. Unfortunately, this criterion does not work for such compounds like some of the examples given in Downing (1977, 826) in which the relation can hardly be considered permanent or habitual: *dinner-bath* (interpreted by a subject as a bath taken in preparation for dinner) and *oil-bowl* (interpreted as the bowl into which the oil in the engine is drained during an oil change).

The fact that compounds have a special meaning of there own is attributed to compounds' semantic and logical unity/wholeness, or 'isolation'.[16] Sweet (1900: 25) points out that N+N

[16] The term 'isolation' was first used by Brugmann in 1891.

compounds are isolated not only from the viewpoint of their meaning (logical isolation) but from the viewpoint of their form (formal isolation). According to Sweet, formal isolation is responsible for the way a compound behaves in a sentence, i.e. the way it is modified (by an article or an adjective or another noun) that is typical of a whole unit.[17] By logical isolation Sweet (1900: 26) means that compounds have a special meaning of their own, i.e. logical isolation is described as a relative independence of the meaning of a compound from the meaning of its constituents. So, the meaning of a compound is not a simple sum of meanings of its elements, but compounds are ultimate sense-units where the meaning of the constituents contributes to the meaning of the whole only to a degree. The differences in meaning in pairs of compounds like *cardboard box* and *jewellery box* and *lemon tree* and *lemon juice* are explained by the isolation of the meaning of the compounded lexemes from the meanings of the constituents comprising them.

Isolation (both logical and formal) is commonly considered responsible for a number of compounds becoming traditional/conventionalised names for single objects and phenomena, which may result in a (partial) loss of semantic transparency. For example, the lexicalised compound *birthday* is not normally used to refer to the exact date of someone's birth, but to denote one day in a year when someone's birth is celebrated. In examples like *cruise missile* the connection of the meaning of the compound, i.e. 'a low-flying missile which is guided to its target by an on-board computer' (Oxford online), to the meanings of the constituents is less clear than in *birthday*. And in some cases, e.g. *holiday*, *honeymoon* and *flash drive* do not seem to be directly motivated by the semantics of the constituents at all.

The considerations above bring us to the questions of lexicalization and semantic transparency, which are discussed in the next section.

2.5 Lexicalization and compositionality of meaning of compounds

2.5.1 Lexicalization

Giegerich (2009), when talking about lexicalization, says that it is a gradient phenomenon and although the notion of lexicalization is intuitively appealing, it is rather poorly defined in literature. Some authors (e.g. Berman 1986, 2009) even suggest that the term 'lexicalized' should be avoided

[17] Bauer (2001: 695) develops the idea of formal isolation and talks about grammatical isolation and phonological isolation. Grammatical isolation is described by him as the impossibility for the elements' taking the full range of inflectional marking, modifiers, auxiliaries and determiners. By phonological isolation he understands specific phonological means that make compounds distinct from syntactic constructions.

altogether, since lexicalization is essentially a matter of degree, and speaker judgments differ in how they rank compound expressions for familiarity (Berman and Ravid 1986).

In Cognitive Linguistics, the notion of lexicalization is strongly tied to the so-called 'degree of entrenchment' of a unit. The term 'entrenchment' is aimed to convey the idea that a lot of what speakers say is based on pre-packaged units that are 'entrenched' in our memory so deeply that their activation is practically automatic (Schmid 2007: 118). If units are entrenched, they have achieved the status of conventional items (Langacker 1990: 45) and their activation becomes automated to the extent that they have been used before. Therefore, constructions fall "along a continuous scale of entrenchment in cognitive organization" (Langacker 1987a: 59).

As noted by Bauer (1983: 50), lexicalization is a diachronic process but the traces it leaves in the form of lexicalized lexemes have to be dealt with in synchronic grammar. In the case with nominal compounds, the process of lexicalization moves a compound away from being pragmatically interpretable to being indivisible lemmas which we might not even be able to analyze on the level of structural representation, as is the case with such extreme cases of lexicalization of (former) compounds like *husband* (from *hús* 'house' + *bóndi* 'occupier and tiller of the soil') or *lord* (from Old English *hlāford* 'loaf-keeper'). In morphologically analysable cases like *honeymoon* and *Sunday* we deal with a conventionalized figurative, obsolete meanings that are no longer available from the meanings of the constituting elements. The meaning of such compounds is known to native speakers and it has a common usage within a speech community. The meaning of lexicalized compounds is often reported to be accessed directly from the lexicon, since they are lexical items and thus should have a lexical concept node in the same way as other lexical items.

According to Brinton and Traugott (2005), there are a number of ways in which an item can be lexicalized but most often lexicalization is connected with phonetic developments. For example, if we look at the pair of words *pan* and *saucepan*, we can see that the former is a monosyllable and has full stress, /'pæn/. The latter, however, has reduced stress on the second syllable, /ˈsɔːspən/. Probably, the phrase *sauce pan* will be understood as 'a pan used for cooking a sauce in', which is not necessarily true for the compound word *saucepan*. So, as a result of lexicalization, the compound has also partially lost semantic motivation and nowadays the conceptual difference between the words *pan* and *saucepan* is not based solely on their sphere of usage or purpose, but on the fact that these two even look different: a pan is typically flat and broad whereas a saucepan is normally considerably deeper. Therefore, we can say that *saucepan* is lexicalized, since at the current state of the language *saucepan* cannot be derived from *sauce* and *pan* by means of

compounding if it retains its form and meaning. Being a lexicalized compound, it is understood as a single word by speakers of English. To sum it up, the main features that lexicalized compounds typically possess are: they are no longer derived by a productive morphological process, their phonological form undergoes changes and they possess a specific meaning of their own[18], which may be relatively independent from the meaning of the constituents.

Based on the brief explanation above, lexicalization is sometimes considered one of the criteria for distinguishing compound words from phrases, i.e. we should expect compounds to be at least partially lexicalized and to be listed.[19] But as Lieber and Štekauer (2009: 7) rightly point out, it does not take long to dismiss lexicalization as a criterion for compoundhood, since where the process of compounding is productive; it cannot be the case that all attested compounds are listed/lexicalized.

Moreover, in many respects lexicalization is a matter of personal judgement and interpretation. For example, such compounds as *birthday* are assumed to be lexicalizations with a specialized meaning, whereas cases like *wood people* (used in reference to people selling and delivering firewood, as in the sentence: *The wood people were supposed to come today*) are creative formations many of which can be constructed on the fly.

It seems common for generative approaches to make a clear-cut distinction between lexicalized and non-lexicalized N+N sequences (e.g. Anderson 1992). However, another problem with using lexicalization as a test for compoundhood is that it is not always possible to uncontroversially label something as being an item of lexicon for at least two reasons, as noted in Plag, Kunter, Lappe and Braun (2008: 764). First, lexicalization is not a categorical notion, but rather a gradual one, and second, it is not clear how it can be decided whether a given item is lexicalized or not, or (if a gradient view is assumed) more lexicalized or less lexicalized than another item. For example, *honeymoon* has nothing to do with either *honey* or *moon*; hence it is a single lexical item. In cases like *lemon tree, health centre, boy band* it is possible to argue that these may be productively composed, despite the fact that all of them are listed in published dictionaries as compound words

[18] It is necessary to point out that there are plenty of places (e.g. *blackmail, white lie*) where there is no phonological isolation but which demonstrate a high degree of semantic specialisation.

[19] Di Sciullo and Williams (1987), who invented the term listedness, point out though that it is not entirely a property of words (compounds). Despite the fact that words are more likely to be listed as compared to phrases or sentences, examples like *a state-of-the-art computer* and *The road to hell is paved with good intentions* serve as evidence that something does not have to be a word in order to be listed. As noted by Bauer (1998: 68), using listedness as a criterion is not just useless for distinguishing between compound words and phrases and may also cause a big problem: if all compounds are listed, then it is difficult to see how compounding can be productive, and yet new N+N compounds are coined on a regular basis.

and should be considered to be single lexical items. However, in a situation when a speaker encounters items like *honeymoon* and items like *lemon tree* for the very first time, the meaning of the latter will be easier to access because of the connection that holds between the meaning of the whole and the meanings of the constituents.

Moreover, current psycholinguistic models of the mental lexicon are based on the assumption that lexical representations that are contained in our memory have various degrees of strength (e.g. Butterworth 1983). What do different degrees of strength depend on? If we look at Lipka's (1994: 2165) definition, according to which lexicalization "is defined as the process by which complex lexemes tend to become a single unit, with a specific content, through frequent use", it is obvious that frequency of usage is an important factor for lexicalization. This means that the more often we encounter (or produce) a certain sequence of words, the easier the access to this sequence is and the more likely it is that this sequence becomes lexicalized. Based on the above the question arises: how often does one need to be exposed to a sequence for a lexical representation to be established? In order to answer this question a number of factors should be considered like the discourse in which the item is likely to appear[20], the importance of the item for a speaking community (or an individual) and even the quality of an individual's memory. As suggested in some recent research (e.g. Bloom 2000, de Vaan, Schreuder and Baayen 2007), even a single exposure is enough to leave detectable traces in memory. However, more frequent exposure is necessary for a lexical representation to be strongly established in long-term memory.

In this research I follow the position that lexicalization must be understood as a continuum that ranges from highly familiar compounds like *bathroom* to novel combinations *beach beverage* rather than a strict dichotomy of lexicalized vs. novel compound coinages (Bauer 1983, Gagné and Spalding 2006b, Nakov 2007, Plag et al. 2008, etc.). This position allows us to view lexicalization as an expected step in the process of word-formation, when a lexical item receives a permanent listing in the mental lexicon (Plag 2003), which, in its turn, allows for a shift or specialization in meaning (Bauer and Huddleston 2002). So, the lexicalized meaning of a lexical item is contrasted with its compositional meaning. This brings us to the question of what is compositionality of meaning of nominal compounds and how it is related to semantic transparency.

[20] Naturally, compounds whose usage is limited to a certain discourse are more likely to be lexicalized for the speaking community within this discourse. For example, *bubble boy* (referring to the person who finished the poker tournament on "the Bubble" - the last place which does not pay any prize money) for a regular pub poker player does not require much semantic processing and can be considered lexicalized in this discourse. However, the same compound would mean something different for a medical specialist (a kind of virus that causes bubble boy disease) or for a person working in IT (a kind of e-mail virus).

2.5.2 Compositionality

In Section 2.4 I looked at such unique characteristic feature of compounds as isolation and discussed the literature on this issue, most of which dates back to the beginning of the 20th century. With the development of Construction Grammar (Fillmore 1972, Croft 2001, Goldberg 1995, 2006, Lakoff 1987, Michaelis 2004, etc.) and lexical semantics (Cruse 1986, Jackendoff 2002, Levin 1993, Pustejovsky 1995, etc.) the term 'compositionality' seems to be favoured by researchers working in the field of formal semantics.

It is necessary to briefly discuss the issue of compositionality as related to nominal compounds. In this section I will look at some of the questions that are most commonly discussed in relation to this issue. First, the underlying idea of compositionality is that the semantic value of the whole stems from the principle of combination. So the meaning of a nominal compound (XY) can be put down to the meaning of X, the meaning of Y and the principle according to which X is related to Y.[21] For example, Granville Hatcher (1960) proposes that most XY compounds describe one of the following four relations between X and Y: 'X is contained in Y', 'Y is contained in X', 'X is the source of Y', or, finally, 'Y is the source of X'. For example, *doghouse* and *house cat* epitomize the first two relations, whereas *sugar cane* and *cane sugar* are representative of the latter two.

The two main approaches that deal with the issue of compositionality are the generative approach and the classificatory approach. The classificatory approach aims at classifying compounds on the basis of the relations between the components. However, it has long been noticed that creating a properly exhaustive semantic classification is rather problematic. Compounding, being a productive pattern of word-formation in English with a large number of items being coined every day, is difficult to organise into nice and neat types and categories. Even with the most precise classification there will still be cases that do not fit into any category. For example, as noted in (Bauer 1978: 58), in the case of *fire alarm* it is difficult to say that fire is the source of the alarm, especially if the alarm is never activated.

In the early generative approach of Lees (1968)[22] the main idea is that a compound is produced by means of deleting a syntactic deep structure, i.e. a compound in itself represents a compressed sentence because of the proposition implied in its structure. The main drawback of this approach is that there is no way to possibly retrieve the one and only right sentence even with semantically

[21] This is usually referred to as 'principle of semantic compositionality', which states that "… the meaning of an expression is the function of, and only of, the meanings of its parts together with the method by which those parts are combined" (Pelletier 1994: 11).

[22] Different approaches to nominal compounds are discussed in Section 2.6.

transparent compounds. In the example with *police dog* it can either be 'a dog used by the police', 'a dog that serves the police', 'a dog trained by the police', 'a dog that wears a police uniform', etc.

Unlike the term 'isolation', which needs further specification, i.e. 'formal isolation' or 'logical isolation', the term 'compositionality' covers both structural and semantic sides. Cruse (2006: 29) explains compositionality as working out the meaning of an expression by combining the meanings of its constituents. This approach to compositionality seems rather simplistic, since, as Construction Grammar points out, constructions carry their own meaning; and a sequence of a preposition and a noun (e.g. *in Wellington*) does not follow the same interpretational pattern as a sequence of a noun and a verb (e.g. *play the piano*). So, the meaning of the construction is as important for its interpretation as the meaning of its constituents and should be accounted for.

In a most general understanding the standard compositional claim for compounds has been formulated as follows (paraphrased from Sweetser 1999: 131-132):

1. The meaning of a complex expression is a function of the meaning of its component elements and the syntactic rule that governs their combination.
2. In compounds, the function of the first element is to modify the semantics of the second element.
3. Each element represents in itself an invariant semantic building block, i.e. when one element is qualified by another, only some of its features are modified or highlighted or some new feature is added to it. The head element itself remains semantically invariant.
4. Compounds are ruled by a compositional principle that allows strong prediction of semantic behavior (as a result of this, one word cannot mean very different things in different constructions).

Although the main principles as outlined above are clear, the question of whether compounds are compositional or not is in dispute. Some approaches (e.g. Guevara and Scalise (2009)) suggest that no compound is compositional, others (e.g. Bauer (1979)) claim that all compounds are compositional. Ackema and Neeleman (2004: 81) take an intermediate position and argue that compounds are noncompositional despite relying on pragmatic information.[23] An interesting approach to the issue of compositionality is presented in Kavka (2009: 28), who regards compositionality as:

[23] Their claim is probably based on Jespersen's (1942: 137) and others' opinion that a unit can be called a compound only if its semantics are unpredictable.

> … the product of the continual complex interplay of variability and literalness, and as a complex phenomenon whose common denominator is the reference to semantic unity.

This definition does not seem to include the formal component, but it becomes clear that the unity of the form (formal isolation) is somehow related to the unity of the sense (logical isolation). Kavka (2009) believes that compounds are noncompositional because they are invariable (the order of the constituents is fixed, attributive determinations are only external and affect the whole structure rather than separate components). Any deviance from this results in a change or even complete loss of meaning. As we can see, by trying to prove that compounds are non-compositional, Kavka (2009) sees compounds as subtype of idiom, which, as has been shown before, is not the case. The difference in the meanings of compound *muffin top*, for example, as in the sentences below, can be attributed to the degree of compositionality one and the same compound may possess depending on the meaning it realises in a given context.

Curious about how many calories are in blueberry muffin top?

If you can't see your toes because your muffin top is in the way, it's time to take action.

Another thing that seems to hold responsibility for the integration of form and meaning is 'literalness'. By literalness Kavka understands that every language unit possesses a basic, fairly concrete, standard meaning. He also believes that literalness can be measured and therefore it is gradable, e.g. the degree of literalness in compounds *playground* and *playboy* is different. According to Kavka (2009: 22-23), in some multi-word expressions each constituent has its basic, fairly concrete, standard meaning, referred to as 'literal'. If the total meaning of the whole expression is understood as a cumulation of individual literal meanings, then the expression will also be literal, e.g., *a beautiful woman*. In examples like *black coffee* we can see a smooth transition toward another category of multiword expressions in which at least one of the constituents does not show its literal meaning. Thus in *a blue film*, for instance, only the constituent *film* carries its literal meaning, while *blue* has nothing to do with the colour it normally denotes. If we come across expressions in which no constituents are used in their literal meanings, such as *spill the beans,* they will unlikely be understood literally, therefore, their meaning is non-literal – or perhaps figurative. Kavka believes that it may be useful to distinguish between the terms 'non-literal' and 'figurative', even though they are synonymous and often used interchangeably. This is because there are expressions which can be interpreted both literally and non-literally, e.g. *grasp the nettle*, and there are expressions that can only be read non-literally,

e.g. *white lie*. Then the interpretation 'grab/catch hold of the nettle (by hand)' will be literal, while 'deal decisively with an unpleasant matter' will be referred to as figurative.

Kavka also claims that all compounds possess figurative meaning, and the degree of 'figurativeness' is also different in different units. The two scales are expected to be the inverses of each other, i.e. the more literal a compound is, the less figurativeness it possesses and vice versa. In Kavka's opinion compositionality must be considered a scalar phenomenon as its core components (variability and literalness) can differ in degree. For example, at one of the scale lie such metaphorical compounds as *ladyfinger* (a pastry that resembles a female finger) and *birdbrain* (a person whose brain supposedly has the size of a bird's brain), which are also highly lexicalized. These are contrasted to such examples as *pot plant* and *bus stop*, which are lexicalized but whose meaning is easily predictable. Examples like *boyfriend*, *healthcare*, *lemon fish* lie in the continuum between the two poles demonstrating different degrees of figurativeness/ literalness as well as lexicalization.

A potential problem with this approach is represented by homonymy since it does not provide an explanation how one should treat potentially ambiguous combinations. Lyons (1978: 535) discusses the example *country house*, which denotes a much more limited type of dwelling than 'a house in the country' implies and is, therefore, noncompositional. At the same time in a sentence like 'I don`t like country houses' the same word combination is ambiguous in terms of its compositionality and meaning (Lyons 1978: 537).

However simple, logical and fairly straightforward, we can see that the claims for compounds being compositional as well as their being noncompositional can be called into question. Real data provide a considerable number of examples that do not seem to fit a chosen framework. However, despite the fact that there are a number of problems with the approaches to the question of compounds' being compositional or not, two things important for the current research are made clear. Firstly, an expression at the language level underspecifies its referent, i.e. the overall meaning of the compounded structure is not equal to the sum of the meanings of its components, but should be understood rather as a certain type of a concoction of meanings. As a result, a given expression potentially can activate very different conceptual meanings depending on the context and the speaker's intention. Secondly, it is quite problematic (if at all possible) to define a rule

according to which words are combined in order to produce new words if we base our assumptions on linguistic information alone and disregard everything else.[24]

The proponents of Cognitive Grammar (e.g. Langacker 1987a, Coulson 2001, Fauconnier 1997, Sweetser 1999, Turner 1996) believe that compositionality, or rather noncompositionality, should not be viewed only as a linguistic issue. Compounds' analysability[25] (which is said to be equivalent to transparency of meaning) is believed to be of more value for the linguistic analysis. Differences in degrees of transparency are closely associated with the corresponding differences in degrees of productivity, i.e. the more transparent, hence the more analysable the structure is, the more productive it is in a language. The cognitive approach does not concentrate on compositionality as such, supposing that all compounds have to be noncompositional at a certain level of representation. The term 'compositionality' in Langacker's (1987a: 449) understanding denotes an objective relationship between the composite structure and its components. This is opposed to analysability, which introduces the psychological perspective of the hearer/reader. As Langacker (1987a: 457) explains, composite structures may be, but are not necessarily analysed by the hearer/reader in the process of communication, and we cannot be absolutely sure how conscious or unconscious this process is.

2.5.3 Semantic transparency and/or compositionality

In the most general understanding, semantic transparency is understood as a degree to which the meaning of word, a compound or an idiom can be predicted from its constituents. Clark (1993) explains that the principle of semantic transparency reveals a one-to-one match between form and meaning, i.e. such that each word in a compound stands for a specific element.

A number of studies provide evidence that the notion of semantic transparency is crucial for understanding how multimorphemic words are represented and accessed in the human mind. Libben (1998) presents a model of compound representation and processing. The central notion of this model is that of semantic transparency. Libben (1998) distinguishes between semantically transparent compounds, whose meaning is predictable from the meaning of the constituents (e.g. *blueberry*), and semantically opaque compounds which, according to Libben (1998), are monomorphemic in the minds of language users (e.g. *strawberry*). In other words, while

[24] The necessity of accounting for extra-linguistic factors was pointed out much earlier (e.g. Lees 1970).
[25] 'Analysability' as referred to by Langacker (1987a: 457) is "… the extent to which speakers are cognisant (at some level of processing) of the contribution that individual component structures make to the composite whole.

strawberry can be analyzed into *straw* and *berry*, the meaning of *strawberry* does not have anything to do with the meaning of *straw*.[26]

Libben (1998) distinguishes two types of semantic transparency: constituency and componentiality. Constituency refers to the use of morphemes in their original/shifted meaning. For example, in *shoehorn,* the modifier is transparent because it is used in its original meaning, while the head is opaque. Componentiality pertains to the meaning of the whole compound, e.g. *moneybags* is non-componential because the meaning of the whole cannot be directly inferred from the meanings of its constituents.

The results of further studies by Libben, Gibson, Yoon and Sandra (2003) demonstrate that semantic transparency sets boundary conditions on whether a complex word can be comprehended in terms of the constituents that comprise this word or whether it must have its own representation in the mental lexicon. In the course of our discussion of lexicalization and compositionality I pointed it out that compound words differ in the strength of the relationship between the meaning of the whole compound and the meaning of its constituents. For some compounds, the meaning of both constituents is directly related to the meaning of the compound word as a whole. One of the examples often used in the literature is the compound *car-wash*, whose meaning can be easily comprehended by someone who had never heard this word before. Therefore, *car-wash* is usually described as semantically transparent because the meaning of the whole sequence can be derived from the meaning of both constituents.[27] In contrast, a word like *hogwash* requires an independent entry in the mental lexicon since its meaning 'nonsense' is idiosyncratic and cannot be predicted from its constituents *hog* and *wash*, i.e. when used in this meaning *hogwash* is semantically opaque. Compounds differ in the type of opacity, since there are compounds for which either the first, the second, or both, constituents can be opaque. Examples like (1) *horseradish, strawberry* and (2) *doughnut, jailbird* are partially opaque, with compounds in (1) having the opaque modifier, and compounds in (2) having the opaque head. Based on Libben's (1998) and Libben, Gibson, Yoon and Sandra's (2003) considerations, Dressler (2005) discusses the scale of morphosemantic transparency. This term is used by Dressler (2005: 271) in order to oppose it to the term "semantic compositionality", which, in his view, can only hold for (non-

[26] It should be noted that there is a folk etymology which re-interprets strawberries as berries around which you put straw to protect them from the ground/frost etc. Another opinion is that it has something to do with the berries ripening on straw to prevent them from rotting on soil.

[27] However, even in the case with compounds that are traditionally analysed as semantically transparent, we can see the elements of meaning that do not seem to be easily predictable from the meanings of the constituents. For example, in the case with the compound *car-wash* the presence of the locative meaning in the interpretation is difficult to explain based on the semantic content of either of the modifier or the head.

idiomatic) syntactic units, as they are usually derivable from the meanings of their constituents. The scale of morphosemantic transparency uses the idea that each compositional word is analyzed through the transparency of its constituents, i.e., the recoverability of their original meaning in the meaning of the compound. A binary value of [+/- transparency] is assigned to the head and to the modifier of a compound and, based on that, it is possible to distinguish four fundamental degrees of morphosemantic transparency of compounds, with type 1 being the most appropriate and type 4 the least appropriate in terms of meaning predictability:

1. transparency of both members of the compound, e.g., *door-bell*;
2. transparency of the head member, opacity of the non-head member (modifier), e.g., *straw-berry*;
3. transparency of the non-head member, opacity of the head member, e.g., *jail-bird*;
4. opacity of both members of the compound: *hum-bug*.

However, this does not seem sufficient, even if it is a necessary step. The main reason why it is insufficient is that there are levels of transparency within each element. For example, X in XY may be more transparent than X in XZ, as is the case in pairs like *health centre* and *health problem*, *wind sculpture* and *wind band* (For discussion see Section 5.4.1.6).

Moreover, as Frisson, Niswander-Klement and Pollatsek (2008: 88) rightly point out, the notion of transparency is relative, and even in the case of transparent compounds their meaning cannot be unambiguously predicted from the meanings of the constituents. Thus, a *woodshed* can either be a shed for storing wood or a shed made out of wood, *a cowboy* is rarely ever a male human who is young enough to literally be a boy. Moreover, as shown by Benczes (2006a, 2006b), creative compounds that are usually considered (partially) semantically opaque, e.g. *muffin top, flame sandwich, hen party, jailbird*, etc. can be systematically analysed in the Cognitive Linguistics framework. In my research I agree with Fusté-Herrmann's (2008) position, that semantic transparency must be viewed as a continuum, where one end reflects a more superficial, literal correspondence and the other end reflects a figurative correspondence.

Current linguistic research that investigates the issue of compounds' transparency is concerned with the role it plays in word recognition and in the organization of the lexicon. There are a number of attempts to model the way in which language users access the meaning of a compound. I will briefly outline the main claims of the models that are most often discussed in the

literature and then will talk about the issue of transparency in relation to the formation of N+N compounds.

The whole-word model (Butterworth 1983) claims that a compound is processed by looking it up as a whole, which means that we should not expect to find any influence of the difference in the degree of transparency of the constituent morphemes.

The morphological decomposition model (e.g. Taft 1981) would require compounds to be first decomposed into their constituents. Then the meaning of each of the constituents should be accessed and put together again. This model predicts that opaque or partially opaque compounds will be more difficult to process since the meanings of the constituents of a compound can be in conflict with the stored whole-word meaning.

Other models (e.g. Baayen, Dijkstra and Schreuder 1997, Laudana and Burani 1995, Marslen-Wilson, Tyler, Waksler and Older 1994, Pollatsek, Hyönä and Bertram 2000, Schreuder and Baayen 1995) assume that both the whole-word lookup route and a morphological decomposition route are available and that semantic transparency plays a role in deciding which will be used in each case.

Though it is widely accepted that 'semantic transparency should play a prominent role in any model of compound representation and processing' (Libben 1998: 35), the evidence for semantic transparency effects is not entirely uncontroversial. It is necessary to point out that the results of most of the recent studies indicate that the meaning of the constituents of opaque compounds might not be available during the processing of compound words, even though the lexical forms are retrieved. In contrast, the constituents of transparent compounds are represented (and available) at both the lexical and conceptual level. In this respect it is necessary to distinguish between lexical and conceptual information in compound word processing, which is done in several theories of compound processing (e.g., Libben 1998; Zwitserlood, Bolwiender and Drews 2005).

Based on the discussion above, we can see that the terms 'compositionality' and 'semantic transparency' are used to refer to similar (if not one and the same) notion by different authors. The main difference seems to be that it is customary to talk about compositionality of meaning of compounds, but semantic transparency of components. This means that semantic transparency is mainly understood as a property of individual constituents comprising a compound rather than of the entire multimorphemic expression, which somewhat differs it from compositionality, where the relation between the components also matters. Dressler's (2005: 271-272) understanding of

semantic transparency sheds some light on the relation between the two notions. According to him, morphosematic transparency of compounds should be viewed as full compositionality. So, compositionality is assumed to be one of the necessary prerequisites of transparency and it should allow speakers to identify the meaning of a compound faster, thus facilitating processing of a compound's meaning. Dressler defines a transparent compound as "... one whose meaning is a subset of the set of potential meanings of the compound as constructed grammatically via the combination of the meanings of the two parts" (Dressler 2005: 271).

How is the issue of semantic transparency relevant for the analysis of the principles of formation of nominal compounds? The current research looks into the issue of compounding and views it as a productive process. Therefore, in my research I am mainly interested in the analysis of the items that are not highly lexicalized, comparatively compositional, and relatively semantically transparent i.e. whose meaning is not fixed and is possible to at least partially predict from the meaning of the constituents.

2.6 Different approaches to compounding

The issue of compounding has been investigated from different perspectives and each of the approaches has brought up some important issues, the analysis of which helps shed light on the nature of this phenomenon. I will give some necessary details in order to outline the framework of the analysis chosen for the current research.

2.6.1 Lees's transformational approach

One of the first really thorough approaches to nominal compounds is that provided by Lees (1968) within the framework of early Chomskyan grammar. Despite its innovations and the insights by the application of the Chomskyan model, a number of problems have been noted subsequently, and the objections to Lees as well as the model Lees provides have influenced later development in the study of compounding.

Lees makes an assumption that nominals[28] in English "are not in themselves sentences, but rather noun-like versions of sentences" (Lees 1968: 54). Therefore, they "incorporate the grammatical forms of different sentence types, and of many internal grammatical relationships within sentences, such as subject-predicate, subject-verb, subject-object, verb object, etc." (Lees 1968: 119). Based on that, it is possible to explain the ambiguity of meaning of compounds if we analyse their deep grammatical structure, which is similar to that of a sentence. This means that compounds carry information in the same way a sentence does, and this information should include a propositional component. At the same time, in compounds, which are noun-like versions of sentences rather than sentences per se, propositions are not explicit. This allows for building a compound so as to address the meaning we intend to convey. Lees (1968) believes that any compound should have just one meaning assigned to it in a given context. The possibility for the variety of meanings that one and the same compound can express results from different deep structures corresponding to different meanings:

Cat milk – the milk for the cat

Cat milk – the milk from the cat

Thus the interpretation of compounds is explained in terms of purely "grammatical knowledge" and the semantics of compounds is put into a complete dependence on grammar. Knowledge of an extralinguistic nature as well as pragmatics are reported to be of no value, which gives grounds for criticisms of this approach. For example, Levi (1978) believes that there are different kinds of

[28] Lees uses the term "nominals" as a superordinate term to refer to different types of nominalizations as well as N+N compounds.

knowledge that influence our usage of complex nominals[29] when we come across them in speech either as speakers or as listeners. It is especially true for novel coinages, in which case, as she puts it:

> … additional steps must be performed by both speaker and hearer in order for the former's creativity (or erudition) not to interfere with successful communication. It is here that independently motivated, and more general, semantic and pragmatic principles must be brought into play by the speech act participants; only in this way will they be able to move beyond the level of generality represented in semantic structure (or, more precisely, in the set of possible semantic structures that they may associate with the surface form) to identify the specific, inevitably complex referent that is named by the new CN (Levi 1978: 238).

Another criticism of Lees (1968) concerns the claim that the transformational approach is applicable when explaining the meaning of cases like *windmill* and *flour mill*. The compositional surface structure of both of them is identical (N+N), but the difference in the underlying structures predetermines the differences in the meaning of these compounds, since in *windmill* – wind powers the mill, whereas in *flour mill* – the mill produces the flour (Lees 1968: 117). The main criticism for this claim is expressed in a number of works, e.g. Allen (1978), Booij (1977), Scalise (1986), etc., who make it the point that while the proposition between the constituents supports the idea of the sentential origin of compounds, the observed semantic regularities can be successfully expressed by other types of rules, for example, rules in the lexicon. And from the lexicalist position (see below) the difference in the meanings of *windmill* and *flour mill* should be encoded in the lexical entries for the given compounds[30].

The next widely criticised point of transformational treatment of compounds is connected with deletion of the lexical material. According to Lees (1968), the proposition should be expressed by a verb in the implied sentence from which a compound is derived. On the way from the deep structure to the surface structure this verb is deleted. However, the problem with deletion of the verb on the way to surface structure is that it is not possible to provide a satisfactory explanation for ambiguous cases like *snake poison* if we try to recover the underlying sentence structure in the

[29] Levi uses the term "complex nominals" (CNs) to refer to three subsets of combinations that share a common structural feature of a head noun preceded by a modifying element. This element in some cases can be defined as a noun and in other cases it appears to be an adjective. According to Levi (1978), the former is represented by the subset of compounds: *doghouse, apple cake* and the two latter are subdivided into nominalizations: *film producer, American attack* and combinations with nonpredicating adjectives as a modifying element: *rural policeman, solar generator* (examples from Levi 1978). This subdivision, though, seems to be very subtle and is in many cases ambiguous. This, as Levi (1978) explains, is the reason why she prefers to use a more general term to refer to the combinations of this kind.

[30] Since the lexicalist approach denies the influence of the syntactic component on the generation of meaning in a compound, the relations between constituents becomes an important question. We will be discussing the issue of semantic relations between the components of compounds in the next section of this chapter.

"surface structure→ deep structure" direction. In examples like that, neither the underlying kernel structure nor the formal surface structure give a sufficient amount of information to identify the meaning of a coinage. Thus, the main problems with this approach are that (a) deletions become dependent on the semantics of the components (which contradicts to Lees's claim that the semantics of compounds should not be used as a basis for the analysis) and (b) deletions become unrestricted. Even if we assume that it is possible to recover the deleted verb when we analyse a compound whose meaning is not ambiguous based on its surface structure, the number of verbs can be immense. As noted by Scalise (1986: 12), it would be necessary to make an explicit reference to what verb is deleted in each single transformation, e.g. for *windmill* – "*power* deletion", for *flourmill* – "*grind* deletion", for *car thief* – "*steal* deletion". These transformations are ad hoc in the sense that each of them is constructed to account for one particular example. This makes a grammatical explanation impossible or, at the very least, unacceptably bulky.

As pointed out by Scalise (1986) the fact that identical underlying sentence structure can produce different compound structures gives more grounds for criticism: e.g. *wading-bird* – 'The bird wades' (the determinatum is represented by the subject of the sentence) vs. *population growth* – 'The population grows' (the determinatum is represented by the verb). Similarly, comparable similar sentence structures can produce absolutely different compounds (*bee sting* – 'a bee stings', *falling star* – 'a star falls'). Scalise (1986: 12) also notes *t*hat the meaning of a compound is somewhat different from the meaning that would be expected based on the deep structure relations. For example, although "a green black-board" is perfectly acceptable, its presumable sentential source, "*a green board which is black", is not, because it is contradictional semantically (not because it is syntactically ill-formed).

Thus, we can see that however plausible and innovative Lees's ideas are, putting the syntactic structure of compounds as the primary purpose of the analysis and attempting to put their semantics into complete dependence on syntax gives grounds for criticism. It is difficult to disagree with Štekauer (2000) who notes that putting an object that needs nomination into the frames of a syntactic structure is a fallacious approach as each syntactic structure that contextualizes a naming unit, makes the meaning of the word undesirably (in terms of naming) concrete.[31] Štekauer (2000: 55) holds the opinion that compounding should be analysed in the frames of word-formation, since the function of the latter is to meet the requirements of a speech-community by coining new naming units for objects, actions, qualities, and circumstances,

[31] However, it is necessary to point out that this was the only way Lees could deal with it, since in the 1960s, for linguists, the only possible way to derive compounds within Generative Theory was by means of transformations.

real or imaginary. This means that users provide linguistic signs for objects (this did not become part of generative theory until the Lexicalist hypothesis several years later). When an object is first encountered, it is the content that is noticed first. The content may include a number of components including shape, colour, purpose, place in the system of extralinguistic relations, personal attitudes, etc. Therefore, the purpose of a linguistic sign is to reflect the content in the first place. All the syntactic information that a sign possesses is relevant to its content only to distinguish between the kinds of things named (thing, process, quality, etc.). So, syntax determines how already existing naming units should be combined into sentences in the process of communication.

Another common reason for criticism of Lees's approach is that Lees seems to be deriving lower-level units from higher-level units. Such practice is considered to be unacceptable from the perspective of functional linguistics (see, for example, Štekauer 2000), as it violates the rules set by the functional approach to language (minor units serve as building blocks for major ones); [32] this, however, is only a problem if one is a functionalist, which Lees was not. At the same time, language provides us with plenty of examples of phrase-based word formation (*to break up – a break-up, to scare crows – a scarecrow*) and clippings (*paper* for *newspaper* and *plane* for *airplane*). Moreover, as is claimed in the systematic approach to language (Halliday 1976, 2002; Berry 1975), units of language of any rank can go one level down and function as units of lower level. For example, the phrase *'I told you about'*, when rankshifted downwards in the NP (as in *'The man I told you about'*), functions as an attribute and is perceived as a united chunk, i.e. 'clause used as a word'. This is also the case for compounds, when the elements constituting a compound are not perceived separately but viewed as an indivisible chunk. At the same time movement in the opposite direction is also possible, the so-called upward rankshift and clippings would be an example of how it works, when one of the components functions as a combination of two components without any damage to the meaning.

So, the problem with the transformational approach is that it uses the principles of syntactic analysis for the analysis of semantics of compounds; therefore, it is quite problematic to use it for solving the problem of ambiguity of compounds, which it is originally aimed at.

For reasons of space I will not go further into detail on the criticisms of Lees's approach here. More detailed discussion on underestimation of the rules in the lexicon can be found in Allen (1978), Booij (1977), Scalise (1986); problems with recoverability of the deleted material are

[32] See Section 2.3 for discussion.

looked into in Booij (1977), den Besten (1983), Marchand (1965, 1974), Scalise (1986); criticism on overlooking the role of pragmatic information can be found in Bauer (1979), Downing (1977), and the importance of extra-linguistic knowledge is discussed in Levi (1978), Ryder (1994), etc. With all these criticisms in mind, there is still no denying that Lees made a considerable contribution to the study of compounding. His work motivated an enormous number of extensive and controversial discussions on a number of issues in word-formation in general, and compounding research in particular. Some of his ideas received reinterpretation in later generativist works (e.g. Aronoff 1976, Levi 1978) and were reconsidered and reanalysed in other approaches, the main tenets of which are outlined below.

2.6.2 The Lexicalist approach and the Variable R condition

The development of the lexicalist hypothesis starts in the early 1970s. Lexicalism proposed a return to the traditional division of the grammar into morphology and syntax. These are held distinct not only because of the nature and size of the units which they concatenate, but also because of the characteristics of the outcomes of such concatenation (Giegerich 2009: 178). In Lexicalism, morphology is integrated into the lexicon and produces members of lexical categories (words) while the outcomes of the syntax are members of phrasal categories. According to Giegerich (2009), reinstatement of the morphological level of the linguistic analysis is dictated by the necessity to account for fundamental differences between syntactic and morphological constructions. The mechanisms of producing phrases and sentences are believed to be productive and the meaning of phrases and sentences is compositional and transparent. According to the Lexicalist position, phrases and sentences are usually created afresh every time we need to pronounce them, but words have a more permanent existence. Words are formed and then (often) retained. If words are retained in a speech community long enough, their meanings and often also their forms are subject to change and may even lose the structural transparency, e.g. words like *lord* and *lady* used to be compounds in Old English. The functions of the lexicon are to be a storage place for words that have already been coined (and may or may not have internal structure), and to be a component of the grammar (the 'morphology'), in which words are assembled from familiar morphological building blocks by means of operations which may or may not be fully productive at the current moment of time (Giegerich 2009: 180).

Lexicalism takes its roots from Chomsky's (1970) *Remarks on Nominalization*, which initiated a different perspective on morphological phenomena, suggesting that at least some complex words (e.g. derived complex words) should be explained as lexical formations rather than as

transformations. This idea was taken on in Halle's (1973) proposal that word-formation should be treated as an autonomous component of grammar because it occurs entirely within the lexicon and is handled by a specific mechanism, which Halle (1973) calls Word Formation Rules. Word Formation Rules are lexical rules that are responsible for the sequential arrangement of the morphemes of a language into actual words. Word Formation Rules can apply to the list of morphemes, forming potential words, i.e. words that are not listed in a dictionary. At the same time Word Formation Rules can refer to other properties of a word. For example, Word Formation Rules can change the lexical category of the input lexical item, as well as the syntactic features associated with it (Scalise and Guevara 2005: 154). These rules have access to the information associated with words but not to the other components of grammar. They are different from other rules of grammar because unlike syntactic and phonological rules, which are necessary for the generation of every single sentence, Word Formation Rules are optional. Aronoff (1976) proposes considering Word Formation Rules as 'once-only rules', meaning that a speaker does not use these rules to form a new complex lexical unit every time they need to use it; once formed it is stored in the lexicon as a fully specified separate item. So, all other information associated with it (semantic, syntactic and phonological) should be stored as well, and this provides an account of how the semantic information is encoded in compounds and why there are differences in meaning of items like *windmill* and *flour mill*.

Since in Lexicalism members of lexical categories are produced in the lexicon and members of phrasal categories are produced in the syntax, there must be a sharp division between lexicon and syntax. However, as pointed out by Giegerich (2009), it is not very clear what the precise formal nature of the lexicon-syntax divide is and whether all of what is traditionally believed to be morphology takes place in the lexicon.[33] Following the position expressed in Giegerich (2009), since Lexicalism is concerned with words, the distinction between units of syntax and units of morphology has to be as sharp as possible.

In the lexicalist tradition compounds are believed to be generated in the lexicon (rather than syntax). Selkirk (1982) (among others, e.g. Williams 1981) proposed lexicalist treatments of word formation that borrowed the form of phrase structure rules from syntactic

[33] Giegerich (2009) provides a very extensive discussion of Lexicalism in relation to morphology in general and compounding in particular. Since the lexicalist hypothesis is mainly concerned with the lexicon-syntax divide, most of the works look at the issue of distinguishing between compound nouns and phrases. As this question is not the priority for the current research, I will only look at the points made in some of the approaches relevant for the discussion, of which Allen's (1978) approach is the most interesting for my research. This by no means implies that other approaches are unsophisticated.

theory. In her account, Selkirk (1982) extends the hierarchical projections of the so-called X-bar schema to apply in morphology. She proposes the hypothesis which entails the claim that word structure has to be accounted for by a context free phrase structure grammar.

Scalise (1986: 90) suggests that the rules for creating compounds (Compounding Rules) are in fact word-formation rules of linear concatenation. Compounds are characterised by two conditions: the Variable R Condition and the 'IS A' Condition as Allen (1978) states it.

The 'IS A' Condition is important for the identification of the head of endocentric compounds in general, and, in our case, N+N constructions, e.g. $[[\text{evening}]_N\ [\text{school}]_N]_N$, both in semantic and categorial terms:

- IS *evening school* A type of *evening* or IS it A type of *school*?
- Given that *evening* is a noun and *school* is a noun: IS *evening school* also A noun?[34]

Since an *evening school* 'IS A' (type of) *school* and it 'IS A' noun, *school* is the head of the compound *evening school*. The 'IS A' Condition posits that: (1) the syntactic category of a compound is determined by its right-hand constituent; (2) the meaning of the whole compound is contained in the meaning of its right-hand constituent. In other words, the set denoted by the compound is a subset of the set denoted by the head.

The Variable R Condition presupposes that new compounds have a range of possible meanings. The Variable R Condition can be connected to the variability of meanings in root compounds, also known as primary compounds. Allen (1978: 91) claims that the Variable R condition establishes a range of possible (as well as impossible) meanings for a given compound and argues that the relation between the components is variable and "may be specified in terms of the interaction of the hierarchies of semantic features of the two nominal elements". An N+N compound can have a variety of meanings based on a variety of relations between the constituents, but what happens is that "usage tends to push the compound towards naming one specific instance of one of these relationships" (Allen 1978: 91). One of the possible meanings can become conventionalised as in the example for possible/impossible meanings for the compound *fireman.*

man who worships fire	* man who contains fire
man who walks on fire	* man who sells fire
man who sets fire	* man who dreams about fire

man who puts out fires (examples from Allen (1980: 10))

[34] This question is obviously more relevant when the constituents belong to different categories, e.g. *high school*.

The Variable R Condition is closely related to or, rather, solely dependent on the semantic content of the first constituent which

> … may fill any of the available slots in the feature hierarchy of the second constituent element, as long as the feature slot to be filled corresponds to one of the features of the filler (Allen 1978: 93).

According to this view, one of the most important factors determining the filling of a slot is the semantic compatibility of the features of the combined nouns. There also exists a hierarchy of semantic features that predetermine which of the slots in the semantic content of the head can be filled. This seems to work fine with cases like *water mill*, which can mean 'mill powered by water', 'mill located near water', etc.; whereas interpretations like 'mill which lives near water' is impossible as the semantics of the second component places certain restrictions as *mill* refers to an inanimate object, and *live* is normally associated with animate ones.[35] Judging by the examples given above, there seems to be some kind of contradiction here, though. On the one hand, the head concept *man* in the above example *fireman* can be characterised as having the feature [+ability to walk], which explains why the interpretation 'man who walks on fire' is acceptable (though very unlikely). On the other hand, it is not clear why [+sell] feature is not given as possible. What can prevent the combination *carpet man* from being interpreted as 'man who sells carpets'? Thus, it looks as though there is more that should be taken into consideration than just the semantic combinability of the constituents. It is also necessary to account for pragmatic information and knowledge of what objects of the extralinguistic reality can be described by a given sequence of nouns.

The meaning of some compounds deviates from what can be expected on the basis of the compositionality principle, e.g. *cartwheel*, *pancake*, *eggplant*, etc. In cases like that the Variable R Condition cannot be applied, and Allen (1978) believes that they should be listed in the lexicon. Verbal-based compounds can be considered an example of structures where this link is not needed since they contain a verbal element which predetermines the semantic relation between the components, e.g. *snow removal* – 'somebody removes the snow', *truck driver* – 'someone drives a truck'.

[35] I have heard, however, examples like 'Where does it live?' in reference to inanimate objects (books, scissors, etc.), e.g. when somebody is helping to sort out things in the house. Such cases are definitely examples of metaphorical extension, but the possibility of such an extension puts into question the claim that some meanings are impossible.

Being strongly opposed to the transformational treatment of compounds, Allen (1978: 114) suggests the Primary Compound Formation Rule, which shows that primary N+N compounds do not have the syntactic origin.

$$[\#X\#]_N \ldots [\#Y\#]_N \rightarrow [[\#X\#][\#Y\#]]$$

Condition: Y contains no V (verb)

Since the external boundaries are not specific, they are not reflected in this rule. Instead, they are assigned by the so-called External Word Boundary Assignment. The Primary Compound Formation Rule forms productive primary compounds on condition that their semantic structure is transparent. By transparency she means "… complete predictability of the range of possible meanings for a given primary compound, assuming the existence of general principles of meaning formation in compounds such as Variable R and IS A Condition" (Allen 1978: 114-115).

However, internal boundaries prescribed by the rule above, the so-called Strong Boundary Condition, suggest that there is no semantic distortion of the components as they do not influence each other. As a result of this, the formation of non-transparent meaning is simply blocked. However, the loss of transparency in semantically transparent compounds is inevitable. We can see how the meaning of one and the same word changes in different compounds. For example, the tree in *apple tree* is inherently different from the tree in *palm tree*, in the same way our understanding of chocolate differs for *chocolate shop* and *chocolate bar*. So, the differences in the mental images of the coinciding elements alter depending on the nearest linguistic environment, because our idea of the specifications of the concepts denoted by these elements.[36] This makes it possible to suggest that semantic distortion does take place; however, it cannot be accounted for in Allen's (1978) framework.

Some of these problems are dealt with in the lexicalist approach of Jackendoff (1975), whose Full Entry Model posits that the lexicon comprises a list of lexical entries ready for insertion in the appropriate contexts, and not a list of morphemes (as suggested by Halle 1973). According to Jackendoff's model, the lexical entries contain the information about the form, meaning and syntactic constraints of the items. This makes the question of the amount of information that lexical entries contain important. Jackendoff (1975: 643) suggests that any lexical entry for a word should consist of the three components: a) the information that the word exists; b) the information about the word that cannot be predicted by a rule; and c) the cost of referring to the

[36] For further discussion of this issue see Section 2.7.4.6.

rule. Jackendoff (1975:655-658) illustrates the application of these three principles using the example of the compound *garbage man*. The fact that this word exists and is listed in the lexicon, as opposed to, for instance, *tissue man*, is accounted for by principle a). The lexicon has an entry for *garbage* and an entry for *man*, which means that all their information will be counted as redundant in evaluating the entry for *garbage man*. If we assume that there is a redundancy rule[37] specifying that two nouns can be put together to form a compound, then the properties of the two nouns are used to determine (many) properties of the resulting unit. There should also be more specific rules giving more specific information about how the coined compound relates to the two nouns. The cost of referring to rules allows for finding the balance between generality and specificity of these rules. The more general a rule R is applicable, the less the cost of referring to R.

Therefore, in Jackendoff's (1975) view, word formation rules have two functions: they function as redundancy rules with respect to existing complex words, and specify how new complex words can be made (Jackendoff 1975). Jackendoff points out, however, that word-formation rules do not tell us much about how the meaning is acquired by a compound and that a speaker's knowledge of the English lexicon is partly responsible for the way in which the meanings of compounds are related to the meanings of their constituents. Thus it is possible to assume that a person simply does not know English if they used the compound *garbage man* to mean 'a man made out of garbage' (by analogy with *snowman*).

The lexicalist approach denies the possibility of a syntactic component for compound creation. However, claims that coining a compound is not at all connected with syntax (Štekauer 2000, Marchand 1965, 1974) can also be called into question. It seems logical to suggest that when a hearer/reader encounters a novel compound, there should be a stage of analysing the combination and unfolding the meaning in order to access the possible semantic interpretation of a compound. Thus, information used for interpreting a compound plays a certain role in retrieving the meaning. When creating a compound to name a certain object we presumably follow not the same, but in many respects a rather similar pattern. At the same time, linguistic information alone does not provide a full picture. In Tarasova (2007) the results of an experimental study are presented, which show that novel combinations like *balloon pregnancy* and *pillow lips* were most often interpreted by the participants as 'difficult pregnancy' and 'swollen lips' (72% and 78%

[37] According to Jackendoff (1975), lexical redundancy rules are used to filter out redundant, i.e. rule-governed aspects of the representation of structurally complex lexical items. These are the rules within the lexical component and they account for the relation between the derivational base and the derivative (in the case of compounding – between the compound word and its constituents). Redundancy rules define the set of all possible compounds of English, and the lexicon lists the actually occurring compounds.

respectively), which demonstrates the influence of background / encyclopaedic knowledge. A somewhat similar phenomenon was noted by Gleitman and Gleitman (1970), whose experimental study clearly demonstrated that the level of education of the participants plays a considerable role in the ability to understand and provide a sufficient interpretation for novel stimuli. Ryder (1994: 122) talks about participants' use of peripheral schemas as opposed to highly predicted central schemas,[38] which she explains by the influence of the background/encyclopaedic knowledge on the interpretation process (see further below).

2.6.3 Levi's (1978) approach to the semantic relations between the components of N+N compounds

It has been noted that in a number of cases endocentric compounds that have a common constituent realise different relations. For example, as pointed out by Bauer (1979: 45) a *sleeping pill* CAUSES somebody to sleep, a *seasickness pill* is used FOR seasickness, and an *antihistamine pill* CONTAINS antihistamines. One of the most often cited classifications of semantic relations is that of Levi (1978). This classification is used as the basis for the semantic analysis of the data in this research. The reasons for choosing this classification over the others (e.g. Brekle 1976, Downing 1977, Granville Hatcher 1960, etc.) are that it seems to be the most influential and that it is quite succinct, relatively precise, and at the same time allows for quite a loose interpretation of semantic relations.

Levi's (1978) classification is presented in Table 2.1. In order to avoid the use of N+N abbreviations the latter are replaced with terms that reflect the relational meaning of the RDPs (Recoverably Deletable Predicates). The numbers 1 and 2 are used to show the directionality of the relation.

Levi's (1978) analysis is probably the most exhaustive account of the semantics of complex nominals[39] within the generativist tradition. Her approach is based on the claim that all complex nominals are derived by either of the two syntactic processes: predicate nominalization and predicate deletion. In the case with predicate deletion, Levi nominates a limited number of highly abstract predicates as CAUSE, HAVE, MAKE, BE, USE, FOR, FROM, IN, ABOUT, and claims that only these predicates may be deleted in the process of transforming a relative clause construction into the complex nominal, e.g. 'cake with apples' → *apple cake*. Although in this research I follow Levi's

[38] Ryder (1994: 122) gives the example of the compound *finger-pin* being interpreted as 'a pin used for finger pricking', which does not seem to be a very obvious interpretation as the idea that a finger may be pricked with a pin is not very central for either of the constituents and is, therefore, referred to as 'a peripheral schema'.

[39] Levi uses the term 'complex nominal' to refer to three sets of expressions: a) N+N compounds such as apple cake, doghouse; b) nominalisations like government invention, city planner; c) Adj+N constructions with nonpredicating adjectives like atomic bomb, musical talent.

classification, I do not support the idea about the deletion of predicates. Following commonly accepted practice (e.g. Gagné 2000), Levi's RDPs are assumed to be semantic relations that hold between the constituents of N+N compounds.

Table 2.1 Levi's (1978) classification of RDPs.

Meaning	RDP	Examples
N1 CAUSE N2	CAUSE1	*Sex scandal, withdrawal symptom*
N2 CAUSE N1	CAUSE2	*Tear gas, shock news*
N1 HAVE N2	POSSESSION1	*Lemon peel, school gate*
N2 HAVE N1	POSSESSION2	*Camera phone, picture book*
N1 MAKE N2	COMPOSITION1	*Snowball*
N2 MAKE N1	COMPOSITION2	*Computer industry, silkworm*
N2 USE N1	INSTRUMENT2	*Steam iron, wind farm*
N2 BE N1	ESSIVE2	*Island state, soldier ant*
N2 IS IN N1	LOCATION2	*Fieldmouse, letter bomb*
N2 IS FOR N1	PURPOSE2	*Arms budget, steak knife*
N2 IS FROM N1	SOURCE1	*Business profit, olive oil*
N2 IS ABOUT N1	TOPIC2	*Tax law, love letter*

Levi (1978: 76) shows that the interpretation of at least three of the semantic relations (CAUSE, HAVE, MAKE) demonstrates bidirectionality, e.g. *tear gas* – 'gas CAUSE tears' vs. *drug deaths* – 'drug(s) CAUSE deaths'. The necessity for distinguishing between the directions of the relations arises from the fact that the relational information is profiled differently and, as a result, the meaning of the compounded structure changes. This change cannot be accounted for if we use, for example, the same CAUSE label for compounds like *heat birth* and *problem children*, since in the first case N1 CAUSES/INDUCES N2, and in the second one N2 CAUSES N1. In the course of this study it was noticed that such oppositions are not limited to the three relations pointed out by Levi (1978: 76). In the collected corpus bidirectionality of the relations is more widespread across the categories and also includes LOCATION (*office building* vs. *garden buildings*), SOURCE (*photon energy* vs. *life energy*), ESSIVE (*mansion house* vs. *tower house*),[40] TIME (*crisis year* vs. *crisis*

[40] The distinction in the direction of the essive relation in compounds seems to be quite subtle and generally follows the principle: ESSIVE1 – every N1 is N2; ESSIVE2 – N2 which is N1. Looking at the above examples, we can see that every mansion is a house, which does not seem to be the case for every tower. However, this principle does not always seem to work. For example, for the compound *building business* the relation is defined as ESSIVE1 because it is possible to say that *building (as industry) is a kind of business*, but *business is building* is somewhat clumsy. The decision on the directionality of the ESSIVE relation was taken based on the more natural interpretation of a compound.

decisions), INSTRUMENT (*farm machinery* vs. *wind farm*).[41] At the same time, it is necessary to point out that usually one of the directions is actively utilised, whereas the cases of the opposite direction can be marginal.[42]

Another modification of Levi's (1978) classification concerns the necessity to distinguish between the SPACE and TIME relations (Levi views TIME as a metaphorical reading of SPACE). The parallelism of the concepts of SPACE and TIME has been long discussed and the overall historical tendency of spatial expressions to develop temporal meanings is recorded across languages. In the cognitive approach this parallelism is explained by Mapping Theory (Lakoff and Johnson 1999, Lakoff 1993, Boroditsky 2000, Radden 2004) according to which the abstract domain of time gets its structure from a more concrete domain of space. In the current research the relations of TIME and SPACE are separated and viewed as two distinct categories. This decision is based on the idea expressed in a number of recent studies. For example, Matlock, Ramscar and Boroditsky (2005) point out that although the domains of space and time share conceptual structure, with frequent use, thinking about time does not necessarily require access to spatial schemas, because mappings between space and time come to be stored in the domain of time. Besides, some studies on locative prepositions in English (Kemmerer 2005, Sandra and Rice 1995) suggest that language users clearly distinguish temporal meanings from spatial ones. The results of Kemmerer's (2005) research suggest that although the spatial and temporal meanings of prepositions are historically linked by virtue of the TIME IS SPACE metaphor, they can be (and may normally be) represented and processed independently of each other in the brains of modern adults. It seems logical to suggest that the concepts of TIME and SPACE are distinguished in the mental lexicon and the same can be said about the corresponding semantic relations.[43]

Levi's (1978) approach rests on the assumption that every complex nominal is semantically ambiguous and this ambiguity originates from the fact that any of the semantic primitives can be deleted in one compound. The ease with which a hearer interprets a novel combination is explained by more or less predictable grammar of complex nominals. She names several strategies that help to disambiguate the meaning, among which are:

1) knowledge of how the units in question are formed;

[41] Plag, Kunter, Lappe and Braun (2008) also note a wider distribution of bidirectionality in the semantic relations, which they call reciprocity of semantic relations.

[42] For example, in the corpus collected for this research, compounds realising CAUSE2 relation occur only 22 times, whereas CAUSE1 is much more frequent, with 244 cases.

[43] The alternative opinion discussed in a number of works on linguistic semantics (Anderson 1971, Clark 1973, Ornstein 1969) views the concept of TIME as a metaphorical extension of SPACE. In this research these concepts are viewed as separate, though their interrelation is not denied.

2) knowledge of naming patterns, which is knowledge of the semantics of the head and the modifier and the way they can be combined;
3) extra-linguistic knowledge (by which she only means encyclopedic knowledge, disregarding conventional usage, personal experience, contextual and situational factors, the knowledge of how the world works) that helps to exclude ambiguity in the process of the meaning identification.

Levi (1978: 84) admits that her approach is relatively general and bears features of abstraction, hence cannot satisfactorily account for all the individual cases, especially the cases of lexicalised compounds.[44] However, it seems that disambiguation of meaning may not be possible even in relatively simple cases like *chocolate bunny*, where at least two RDPs are possible: 'bunny that *is* chocolate' and 'bunny that is *made* of chocolate' (example from Ryder 1994: 28).

In the light of her theoretical position in Generative Semantics Levi (1978) also attempts to link syntactic and semantic components in the formation of complex nominals. She notes that when used together with the semantic and syntactic constraints on the formation of complex nominals, the constraints basically function as some kind of aid for the listener to be able to choose from all grammatically possible semantic structures the single reading that would be most plausible in the given context and to decide what real world object could be named by a given form. So, grammatical knowledge alone is not enough for both coining and understanding complex nominal structures. Apart from defining/picking out the referent for a given lexical unit (complex nominal structure in our case), the meaning decoding process also involves associating the surface form with its 'institutionalised' (Bauer 1983, Matthews 1974) meaning. That is the knowledge of what semantic structure would be most commonly associated with the surface structure of a compound. As Levi (1978: 239) points out, this kind of knowledge is rather sensitive to the changes of the extralinguistic environment we live in. Obviously, things that are now associated with the compound *mouse potato* will definitely be different from those that could have existed (if they did) before the Internet was invented and spread.

[44] This is usually the point of major criticism of Levi's approach. Spencer (1997), for example, says that Levi's primitives are rather general, which allows for a certain degree of "elasticity" of interpretation. As a result it is sometimes hard to know what would count as a counterexample. At the same time, I believe that Levi managed to point out the most common types of interpretation of complex nominals that we may expect given the way language users conceptualize the world.

2.6.4 Nominal compounds and Cognitive Grammar

2.6.4.1 Some basic notions

Since the current research is concerned with the way compounds are formed and how their meaning is construed, the principles of the Cognitive Grammar approach may be helpful. A brief outline of the principles valuable for my study and the reasons why they are valuable are discussed below.

The main tenets of Cognitive Linguistics are connected with objective extralinguistic reality and understanding of this reality by individuals. Language is viewed as a device, a mechanism, which allows for binding the two together. Since a language cannot exist and develop without being used by a human, then humans should be considered in the analysis too. Thus, meaning conveyed by a language unit acquires one of the most central roles in the cognitive approach. Summing up Lakoff and Johnson's (1999) characterization of the traditional, objectivist approach to 'meaning', three types of meanings can be distinguished: meaning as an abstract category, the meaning of a word and the meaning of a sentence. Meaning as an abstract category is mind-independent, objective and publicly accessible. The meaning of a word correlates with things that exist in objective reality. A sentence is true if the words fit the propositions that make up the sentence. However, it is necessary to point out that Cognitive Linguistics claims that meaning can only be objective outside the human mind (at the same time, there is no use talking about meaning from a linguistic perspective if the human mind is not involved), so it is impossible to talk about the meaning of something if there is no one to perceive this meaning. In Cognitive Linguistics meaning equals conceptualization, which implies that the analysis of meaning is not possible without accounting for abstract entities like thoughts and concepts.[45] Conceptualization is one of the central notions in Cognitive Linguistics and is basically any type of mental experience. Langacker explains: "… linguistic meaning is seen as the product of mental activity on the part of physically embodied, socio-culturally grounded human minds" (Langacker 1990: 26). Martinovski (1995) notes that in general, the assumptions behind different approaches within the Cognitive Linguistics framework are related to the ontology and the epistemology of human language. However, the ontological and epistemological notions that cognitive linguists are working with are not always clearly expressed. The reason is that, for example, for Langacker the cognitive theory of language analyzes meaning only on the conceptual level. At the same time, the lack of overt relation to the

[45] Cognitive linguistics denies the notion of pure syntactic form, as grammatical structures motivate and reflect meaning and can hardly be claimed to exhibit form without any content.

ontological and epistemological notions may lead to misunderstanding. For example, Langacker's suggestion that meaning equals conceptualization (mental experiences) may be interpreted in such a way that perception is part of the process of conceptualization. If this is the case, then boundaries between perception and interpretation are somewhat fuzzy. However, Martinovski (1995) points out that if perception is incorporated in the conceptualization process and if we accept that what we perceive is not always exactly what is going on in the extra-linguistic reality (e.g. our perception of colours), then one can see that Langacker's theory deals with epistemology, since he is concentrated only on the process of conceptualization, i.e. we cannot say anything about how the world really is but how we conceptualize it.

Cognitive Linguistics attempts to give meaning a central position in the architecture of grammar. Due to the development of interest in the research of Mental Lexicon, the analysis of meaning shifted to lexical items and hence to the domain of morphology.[46] Ungerer (2007: 651), for example, points out that "the semanticization of word-formation analysis" should be viewed as the most decisive asset for Cognitive Linguistics to stimulate further research. Unlike formal semantics, meaning in Cognitive Linguistics is not a truth-functional type of meaning. Geeraerts and Cuyckens (2007: 14) note that the conception of meaning in Cognitive Linguistics is fully contextualised and anthropocentric:

> [Conceptualisation] involves imagery in the broadest sense of the word: ways of making sense of imposing meaning. Also, the conceptualisations that are expressed in the language have an experiential basis, that is, they link up with the way in which human beings experience reality, both culturally and physiologically (Geeraerts and Cuyckens 2007: 14).

However, although conceptualization is viewed as a purely subjective phenomenon in Cognitive Linguistics, the possibility of scientific analysis of it, and language in general, is not excluded. As Langacker (1990: 26) believes, conceptualisation has structure and organisation. It is necessary to point out though, that the 'subjectivist' stance of Cognitive Linguistics means that we have no direct access to an 'objective' reality; instead, we construe that reality and that construal is determined by what we are as human beings, with our particular bodily apparatus, cognitive makeup and socio-cultural experience. A lot of this is shared by people, and is thus not 'subjective' in the lay sense of the word. Language and language use reflect our construal of reality, a construal that we share with others of our linguistic community. Therefore, this conception of language and language use does not preclude scientific analysis.

[46] The smallest meaningful unit in the language is the morpheme, and morphemes build words, including compounds, so that morphology is central to the meaning of lexical items.

2.6.4.2 Components of Cognitive Linguistic analysis

I will now briefly look at some of the principles of Cognitive Linguistic analysis. One of the grounding statements of Cognitive Linguistics is that language cannot exist outside the human brain and the language system is put into dependence on the language user's cognitive abilities, most important of which (in relation to language) are symbolization, composition, comparison/ categorization and schematization (Heyvaert 2009). Since these cognitive processes are claimed to be important for language use, it is also often suggested that they are reflected in the language system itself. Therefore, in cognitive linguistics, any analysis of linguistic structure must have "psychological reality" (Langacker 1987a: 42), which means that it has to express what we know of cognitive processing.

In Langacker's (1987a, 1987b) Cognitive Grammar approach language and all structures in a language, i.e. lexical, phonological, morphological and syntactic, are defined as a structured inventory of symbolic units. A symbolic unit consists of two poles: a semantic pole (e.g., [MOTHER]) and a phonological pole (e.g., [ˈmʌðə])[47]. Symbolic units come in various sizes and levels of abstraction/schematization. Symbolization, as Langacker (1999: 94) claims, is based on the psychological ability of association, where "one kind of experience is able to evoke another" (Langacker 1999: 94). Symbolization in language is about construing a relation between a phonological structure and a semantic structure evoked by it. Once such a symbolic relationship between the structures becomes automatic, a symbolic unit is formed. That is the symbolic unit *table* has the form [[TABLE]/[ˈteɪb(ə)l]] (Langacker 1987a: 77). Langacker (1987a: 58) argues that the simplest kind of symbolic unit is the morpheme, as in a morpheme "a semantic and a phonological structure participate as unanalyzable wholes in a symbolic relationship". Grammatical constructions are understood as composite symbolic structures and are viewed as results of the syntagmatic integration of basic symbolic units (Langacker 1987a: 277).

2.6.4.3 What happens when units are combined

Semantic units combine with each other, thus producing more complex structures through the establishment of the so-called "valence relation". The formation of a valence relation is only possible when two of the combined units share certain elements of their semantic structure and can thus be unified (Taylor 2002: 229). Taylor (2002), following Langacker (1999) distinguishes three types of relations that can be characterized as valence relations: head−complement (where the complement elaborates a salient substructure of its head), head−modifier (where the salient

[47] Capitalization is usually used for semantic structure, and the phonological structure is represented phonetically.

substructure of the modifier is elaborated by the head) or the appositional relation (where the elements of a structure designate one and the same entity). Understanding of the head of a construction or a "profile determinant"[48] (Langacker 2000) in Cognitive Linguistics, is not very different from the traditional notion of the head (at least in clear-cut cases). When symbolic structures (e.g. morphemes, words) are combined, the composite semantic structure (meaning of the construction) usually inherits its profile from one of its components. That semantic component is schematic for the composite structure, and is called the profile determinant of the structure. The profile determinant is schematic for the composite semantic structure. For example, in the compound *picture frame*, FRAME is the head because the profiled element of PICTURE FRAME is the same as that of FRAME, i.e., a picture frame is a frame and not a picture (compare with the Lexicalist "IS A Condition").

However, the profile determinant is not obligatory in a construction, nor is there a limit of one profile determinant. There may well be cases when the profile determinant is absent from a composite semantic structure altogether, which is the case for compounds like *spitfire* (in the meaning 'a person with fierce temper'). In compounds like *doctor-poet* it makes sense to say that both components are profile determinant. It is also often the case that the level of schematicity between the component that is more head-like and the whole structure, e.g. in *French toast* the overall designatum is closer to TOAST, at the same time, *French toast* is not representative of the category.

The problem with the notion of the profile determinant is connected with the asymmetry in the distribution of the semantic weight between the profile determinant element and the other element of a construction. It is often the case that the profile determinant element contributes little besides the profiling, and the majority of the semantic weight is contributed by the other element. This asymmetry is most clearly seen in derivatives. For example, in assignment is the derivational base (*assign*) or the suffix (*-ment*) the head? The whole complex structure designates a THING (abstractly defined — see Langacker 1987b), and this information is provided by the nominalising suffix (*-ment*). Being derived from the verb *assign*, the noun invokes the process profiled by the verb as its own conceptual base. On this base it imposes its own profile, i.e. the

[48] Susan Lindner (1981) in her dissertation claims the following: semantic structures are generally expected to consist in a designatum which "stands out in bas-relief" against a background of related cognitive structures which are collectively referred to as the base of the particular semantic structure. This designatum is called the profile, and is said to be "profiled" against the base. According to Langacker (1987:183), the semantic value of an expression resides in neither the base nor the profile alone, but in their combination. In other words, the profile determinant is the element that determines the meaning of the construction it participates in.

thing or person assigned for doing something (see Langacker 2003 for more detailed discussion of examples). At the same time, the derivational base *assign* contributes the vast majority of the semantic specifications and in this sense it should be considered a base element (for the discussion on the issue of headedness see Bauer 1990). A somewhat similar asymmetry in the distribution of the semantic weight is noticed to exist in the case of nominal endocentric constructions in general and N+N compounds in particular. For example, Bundgaard, Østergaard and Stjernfelt (2006) demonstrate how in such examples as *child-safe* vs. *shark-safe* the semantic behaviour of the righthand element (presumably the head) *safe* does not remain semantically invariant through the compounds. We can see that in these two examples the head *safe* takes on different significations in these sequences: in the former case it reads 'safe *for* somebody'; whereas in latter case it reads 'safe *from* somebody'.[49] Weiskopf (2007) discusses the examples *axe murderer* and *child murderer*, which have a parallel surface structure, but in which the head element is profiled differently under the influence of the first noun.

2.6.4.4 Meaning construal in conceptual combinations

Another important point regarding the cognitive approach to language is the necessity of semantic adjustment of the combined components. Langacker (1987a: 75-76) calls this process 'accommodation' of the components. The idea of accommodation is taken on and further developed in cognitive models (e.g. Ryder 1994). Both of them state that native speakers combine their knowledge about the constituents of a compound to arrive at an integrated representation of its meaning. It is also typical of cognitive models (Ryder 1994, Coulson 2001) to claim that the semantic relations that are assumed to connect constituents of compounds cannot exist independently of their arguments but they rather emerge from properties of those arguments when the concepts are combined.

In order to create an integrated representation for a given compound a speaker/ hearer/ reader needs to access the schematic representations of the events or situations in which each of the constituents may typically play a role. For example, the meaning construal for the compound *butter knife* requires accessing the knowledge that we have about both knives and butter[50] (i.e., knives are associated with cutting and are often used for the purpose of preparing food; butter is a kind of food and can be cut). If the context does not suggest otherwise, the representation of the

[49] See Section 2.4.7.6.

[50] This is an important distinction between Cognitive Linguistics and Lexicalist approaches, i.e. the fact that CL treats word meaning as encyclopaedic rather than dictionary-like. A word can activate a lot more in a spreading activation network than is included in a dictionary-like view of the mental lexicon.

butter knife (whatever it may be) should be different from a generic representation of a knife, because it will be associated with the event frame of cutting butter. We can also expect this representation to reflect the speaker's/ hearer's/ reader's thoughts and ideas about what a proper butter knife should look like. This representation can then be then put into the frame corresponding to the current sentence or discourse. If it is necessary, the meaning of a compound can be additionally "decorated" with other beliefs and ideas that may be particularly relevant for the situation. For example, in Downing's (1977) study, one of the subjects extended the basic relational semantics of containment of the compound *oil-bowl* and suggested a whole context in which such an object might be used (i.e., for collecting the oil from the engine of a car).

The distinguishing feature of all cognitive models of compounding is that they are usage-based, i.e. they usually assume that patterns noted in meanings of nominal compounds depend on patterns of how language users experience and conceptualise their environment, rather than on some finite set of rules. For example, the data collected for this research suggests that the most frequent relations in the analysed sequences are those of PURPOSE (*test animal*), POSSESSION (*community power*) and LOCATION (*family atmosphere*) relations.[51] The reason for this may be that our knowledge about entities and our experience with them in general suggests that they exist for certain purposes, they may possess something or be possessed by someone and they can be located somewhere. This knowledge also allows for a distinction between entities, i.e. the differences in the purposes of use between the compounds *test animal* and *transport animal* (examples from the BNC) suggest that the entities themselves are most probably (but not necessarily) different.[52]

2.6.4.5 Ryder's schema-based approach

In the model provided by Ryder (1994) regularities in production and interpretation of compounds are assumed to be influenced by compounds that a speaker or reader/hearer has encountered previously. That language knowledge is usage-based is another important characteristic of the Cognitive Linguistic approach. On the basis of usage the inventory of symbolic units grows horizontally through extension (i.e. we coin new units through analogy with old ones), and vertically through schematization (i.e. we induce more abstract templates from the lookalike instances). These templates or schemata are also symbolic units, however, just more abstract

[51] More than 80 % of the compound data receive these interpretations.

[52] It is obvious that nothing prevents a horse or a donkey from being used for medical or scientific tests, and surely one could easily test race horses for speed/stamina and donkeys for stubbornness. Mice or rats can well be used as transport for fleas. However, a situation where this would be justifiable is unlikely to be considered a typical one.

ones. The usage-based nature of language development is important for this research as it can help in understanding how newly coined compounds acquire their meaning.

Ryder (1994) proposes the use of frames (she calls them 'linguistic templates'), which she believes a speaker operates with. She argues that speakers possess a certain number of constructions for established compounds, i.e. expressions that have been accepted by the speech community but whose meaning may change over time. This stock of constructions allows for the development of certain types of schemas[53] that specify how such constructions acquire their meaning. Ryder (1994, 68-69) distinguishes between three types of schemas that participate in meaning construal of N+N sequences: event schema (sequences of events in a particular context, i.e. scripts), entity schema (general characteristics) and feature schema (actions, relations and qualities that all event and entity schemas share, for example, entity schemas like BALL, ORANGE, WHEEL, etc. share the property of roundedness).

Ryder (1994) also suggests that speakers and hearers/readers are able generalise abstract templates relating constituents to probable meanings. These templates may vary in terms of the degree of abstractness and generality. Some templates are very specific and require sharing a common first or second element (a core word[54]), e.g. *X + box = a box intended to contain/store X* (*bread box, cash box, coin-box, lunch box, sandwich box, tool box*) or sea + X = a metaphorical extension of X that is found in the sea (*sea lion*, *sea bed*, *sea anemone*). Whereas others are very general, e.g. s*ubstance that can be contained + container = Container to hold/ store the substance* (a pattern seen in many compounds with different component nouns: *cigar box, coke bottle, grocery bag).* The most general templates appear to be most reliable and bear some resemblance to the semantic relations reflected in Levi's (1978) semantic primitives. Thus, on the one hand, the fact that cognitive models are so flexible is their main advantage, at the same time their suggestions are often believed to be unfalsifiable, since they allow too much room for interpretation instead of providing rigid frames for the linguistic analysis.

[53] Ungerer (2007: 997) points out that the cognitive notion of 'schema' is usually broadly defined as a "superordinate concept, one which specifies the basic outline common to several, or many, more specific concepts". Tuggy (2007: 120) defines a schema as "a flexible way of generalization that is not understood as a fixed a priori rule, but takes account of salience based on frequency of use".

[54] Ryder assumes that some words should be referred to so-called "core words", i.e. words that participate in the formation of a large number of compounds and that consistently realise one and the same meaning within a given paradigm (e.g. the meaning of container intended for storing something for the noun *box* when it occurs in the head position). She argues that core words can be considered evidence for linguistic templates and says that "once a pattern begins, the more established forms there are in that pattern, the more likely it is that a new form will be based on one or more of these forms, or on a slightly more abstract template based on them" (Ryder 1994: 80).

Overall, in cognitive theories, an emphasis is placed on the creation of meaning by both the coiner/speaker and hearer/reader. In her approach Ryder seems to be influenced by Allen's (1978) Lexicalist framework in terms of the importance she gives to the head noun in the process of coining an endocentric N+N compound. According to Ryder's approach, in the process of coining a new item, first, the head (or, in cognitive terms, the profile determinant) needs to be selected.[55] The priority of the choice of the head noun at this stage is determined by the fact that the head element serves as the referential basis for the expression, i.e. a *rose pole* is a *pole*, not a *rose*; and a *squat rack* is a kind of a *rack*, not a *squat*. The next step is the choice of the modifier, which is selected based on the schemas that the modifier shares with the head. For example, if we coin such a compound as *heel cream*, we may assume that the most plausible common element that a hearer will find is a schema involving creams used to apply to one's heels. If the speaker fails to find nouns whose schemas match the schema of the intended referent a creative expression that utilises metaphor or metonymy (e.g. *muffin top*, *jigsaw activity*) may be produced. Understanding of such an expression will require modification of the schemas that a hearer has for the constituents of an expression.

It is obvious that the hearer/reader cannot possibly have full information as to what strategies the speaker (coiner) could have used in the process of forming a novel item, i.e. whether they based a novel item on the already existing template and if so, which one. Apparently, in order to understand a novel item or expression, the hearer/reader has to turn to the linguistic templates (based on their own stock of established compounds) that are stored in their own memory or to templates that seem to possess similar features to the new coinage. If many templates are possible, the hearer will most probably check what information the context can give about the compound's meaning. This means that we should also expect a certain degree of reliance on the discourse context, which may be helpful in disambiguating the meaning of a novel item and choosing a template with a suitable schema. For example, a nonce item **mouse person* can receive a number of interpretations, some of which can be rather predictable, e.g. 'a person that likes mice' (based on the schema *pet + pet owner/lover*, e.g. *cat person*, *dog person*); whereas some of which may not be as easily predictable 'a person that looks like a mouse', 'a person that is afraid of mice', 'a person that wants to have a mouse', etc. However, the discourse context may make less predictable meanings available. The information provided by the template is then checked against what the hearer knows about the way the real world works.

[55] Ryder (1994) does not look into exocentric compounds, as according to her, exocentric compounds lack a profile determinant.

Looking back at the question of 'accommodation' of the components, it is necessary to point out that Ryder (1994) was also the first to look at the notion of valence[56] as applied to N+N compounds from the schematic perspective. In her opinion in order to coin and interpret a compound, it is necessary to find a correspondence between a schema that is connected with each of the structures represented by the components comprising it (1994: 72). A compound is presumed to have a common or identical schema as part of its semantics similarly to valence relations that can be found in a sentence. For example, in the sentence *The boy walked*, the central schema for *walk* presupposes a slot for an entity that can do the walking, and the noun *boy* is connected to the schema *walk*. In another example which Ryder (1994: 72) gives, i.e. *The fish is walking*[57], the schema of the word *walk* does not allow a fish to be the entity that does the walking. Similarly, from the perspective of the component *fish*, there is also no connection to a schema that would involve *walking*. In the case of the occurrences, a language user needs to look for a schema that is similar to the schema for the other component. In this case the values in one or both of the schemas have to be altered in the process of construing the meaning of an expression. It can be that the *walking* schema has to be adapted to the idea of movement for which no legs are required; or the schema of *fish* is accommodated into that of a fish with legs.

Ryder suggests that the same thing can be expected to happen in the creation and interpretation of compounds, even though valence relations between nouns are much more complex than are those between predicates and nouns.

Ryder's (1994) hypotheses are confirmed by the results of her psycholinguistic experiments in which subjects had to provide a definition for nonce N+N compounds. The analysis of the definitions obtained in the course of the study demonstrated that, when presented with a combination the constituents of which do not seem to be related, language users either look for a schema that is common to both of the constituents or try to adjust the schemas of one or both of the constituents of a compound schemas to make them compatible. The availability of a large number of schemas for nouns in general explains why a non-lexicalized compound can be interpreted in a number of ways. Ryder concludes that the semantic relationships in noun-noun compounds (especially in unfamiliar ones) may appear to be chaotic because they cannot be explained by means of a single, general level of semantic relationships. However, that chaos can

[56] Syntagmatic relationships within composite expressions in general.

[57] Although extralinguistic reality can provide us with examples that seem to contradict Ryder's claim, e.g. a number of the so-called walking fish (mudskippers, the Climbing gourami, Walking Catfish, and other amphibious fish) the idea of walking in this case is not as conventional in the case of the fish.

be ordered if one starts the analysis from the hearer and their attempts to link specific lexical compounds to a range of templates.

Ryder's approach to N+N compounds is beneficial for the current research and some of the ideas are developed further in the current study. The key concepts that are explored further here are those of semantic valence and accommodation of the semantics of concepts constituting an N+N compound.

2.6.4.6 Conceptual Integration theory: Reconsidering the issue of compositionality

The Conceptual Integration theory originally developed by Fauconnier and Turner is aimed at extending and complementing the theory of conceptual metaphor (Lakoff & Johnson 1980), but gradually it has come to cover a much wider range of data. As Fauconnier and Turner (1995, 1998, 2002) claim, their theory is closely connected with blending, but not on the level of linguistic representation, that is coining a new lexical item, e.g. *luggage* + *baggage* = *buggage*. Conceptual blending is a general cognitive operation, which is understood as a process of conceptual mapping and integration that pervades human thought.

The core idea of the theory is that the way we think depends on our ability to find analogies between different objects and phenomena (Fauconnier 1997: 11). Communication is viewed from an online perspective, and the language user's ability to evoke short-term memory packages of knowledge, or mental spaces, while producing or interpreting language is given centre stage. A mental space is a small short-term memory conceptual package assembled as we think and talk for purposes of local understanding and action. Mental spaces contain elements and are structured by frames and cognitive models. They are interconnected and can be modified as thought and discourse unfold (Fauconnier & Turner 1998). Mental spaces are different from domains: domains are broader and encompass "many aspects of an experience that are conceptualized as associated" (Cienki 2007: 181), while mental spaces are more specific in that they involve conceptualizations by the individual in a given context and for a given purpose (for examples of (types of) mental spaces, see Fauconnier 2007: 13-18). A mental space network connects an array of mental spaces that may be linked up with schematic knowledge (e.g. the frame of 'walking') and with specific knowledge (e.g. a memory of the time you climbed Mount Rainier in 2001) (Fauconnier 2007: 351). The mental space that includes the latter experience (i.e., you, Mount Rainier, the year 2001, and the process of your climbing the mountain) can be activated on a number of different occasions.

The integration of two or more mental spaces within a construction is called 'conceptual blending' and a ‘conceptual integration network’ is a mental space network that contains one or more blended mental spaces. A blended mental space is an integrated space that receives input projections from other mental spaces in the network and develops emergent structure not available from the inputs.[58] Conceptual blending operates under a set of constitutive principles and a set of governing principles and involves at least four mental spaces (Taylor 2002: 530): at least two input spaces; at least one generic space, which facilitates the establishment of correspondences between elements of the input spaces; and, a blended space, which inherits selected elements from the input spaces, as well as has emergent meaning of its own.

One of the questions that nominal compounds prompt is how the relational meaning between the elements arises. This relationship does not seem (as we have seen in Section 2.4) to be derivative from the semantics of the individual constituents. Conceptual Integration Theory, with its notion of signification emerging from bringing together of meaningful elements, looks like a promising avenue of exploration for an explanation of this phenomenon. Benczes (2009) utilizes this idea and demonstrates how Conceptual Integration theory can be applied to metaphorical and metonymical compounds, e.g. *fireguard* (an object that prevents the spread of fire), *acid head* (an LSD user), *knee-mail* (prayer), etc.

We can see how the model works by applying the principles of conceptual integration to a combination like *sausage fingers*, for example. In this case the process of meaning identification can be represented as follows: first of all in the process of conceptual integration in the mind of an individual a special integrated network is built on the basis of input spaces. In the analyzed case the input spaces are represented by the notions denoted by the words *sausage* (Input space 1) and *fingers* (Input space 2).

[58] Fauconnier and Turner (1995: 183) call this feature ‘emergent signification’, which is aimed at explaining the fact that the blending of mental spaces may give rise to new meanings that cannot be inferred from either of the mapped domains. This attention to emergent features is what differs Fauconnier and Turner’s framework from the standard metaphorical/metonymical approach (Lakoff and Johnson 1980).

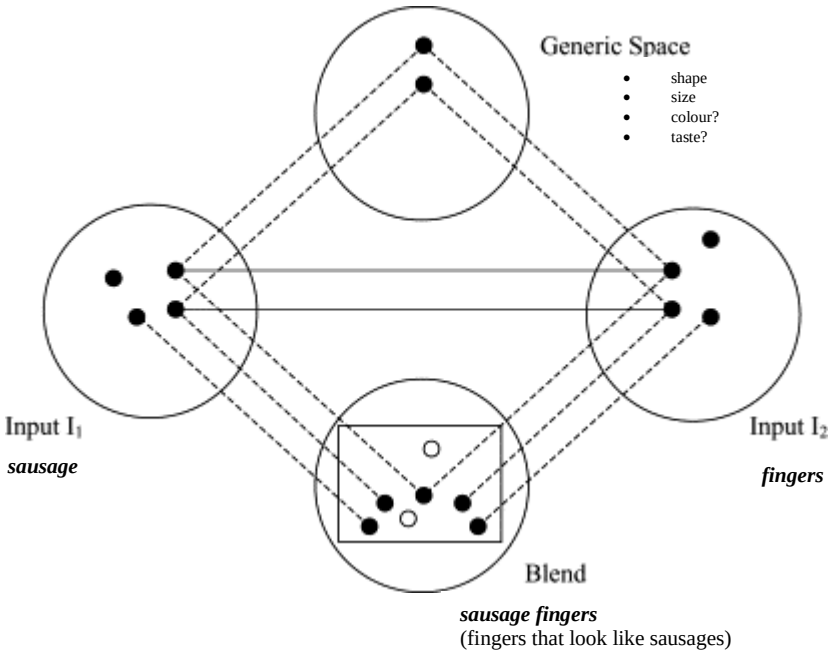

Figure 2.2 The blend analysis of *sausage fingers*

So the associations with the concepts SAUSAGE and FINGERS would probably include the visual image (including shape, size and colour), taste, possible associations based on linguistic experience (expressions where the given word occurs) and extralinguistic experience (personal encounters with both given concepts). As a result of the interaction and conceptual projection there appears a generic space. This generic space is the mental area where the process of choosing properties that could be mapped from the mental representation of the modifier noun to the one of the head noun takes place. In other words possible contiguity points in the combined concepts are searched. This process involves comparing the input spaces as well as identifying the place in the mental representation of the head noun, where the chosen property can be mapped. These processes are really complex. A property of one concept cannot be simply copied and added to another concept because a property and its exact realization depends on or, rather, is limited by the mental representation of another concept and properties typical of it. Nevertheless, we can suggest, that most probably the visual properties of the modifier *sausage* (size and/or shape) could be considered coinciding features with the head *fingers* and can, therefore, be mapped to the mental area represented by the head noun.

In the example *sausage fingers* discussed above, the modifier *sausage* has the meaning of "a short cylindrical tube of minced pork, beef, or other meat encased in a skin, typically sold raw to be grilled or fried before eating" (Oxford online) in the English language, and is only used figuratively either as a component of the old-fashioned expression '*Not a sausage!'* (nothing at all), as an affectionate form of address, especially to a child: *'You silly little sausage'* or when somebody

gives the description of a similar shape (*sausage dog*). In the analysed example *sausage fingers,* the first things we would probably come up with when we think about what the head noun *finger(s)* can denote are what fingers may look and the functions they perform. And here we can see a similar feature, i.e. the shape, which would probably go into the generic space. After further elaboration the new space, the blended space, appears. This already represents in itself a structure which has the basic features of the generic space and also contains some specific elements impossible for the input spaces. In the course of blending the composition operation of the input spaces appears to be the basis for creating the relations which are not typical for the input spaces. The operation of completion recruits background conceptual structure and in the elaboration operation the blended structure gets its final appearance. That is 'fingers that look like sausages'. Background conceptual structures would probably contribute to the connotative meaning, giving the idea of the fingers being fat or short or whatever the personal perception of a sausage might be.

As we can see from the discussion above, one of the basic assumptions and also aims of this approach is that compounds are strictly non-compositional and that the relational information emerges in the process of conceptual integration. However, it is difficult to apply this approach non-trivially to nominal compounds (as a linguistic phenomenon) in all cases. Fauconnier & Turner (1995, 1998) claim that the overwhelming empirical evidence motivates the notion that the semantic behaviour of compounds cannot be predicted, as was shown in the case with *hand wound* and *gun wound*, in which our understanding of the specifications of the element 'wound' changes in each of the examples and the character of this change is not always predictable. Besides, one and the same composite expression may take on novel significations in varying contexts. For example a *duck foot*, apart from its conventional referent, may also refer (a) to a certain type of a weapon (a pistol with several barrels that resembles the splayed shape of the webbed foot of a duck), (b) a leaf of a cannabis plant, (c) a plumbing joint, (d) a kind of a push pole used by fishermen and hunters, (e) footwear that imitates the shape of a duck foot. Bundgaard, Stjernfelt & Østergaard (2006) provide the example of the compound *fire station*, which might refer to different types of stations, depending on the culture and context. In Palaeolithic times, a *fire station* could have been a public place where fire was kept burning so that people could get an ember if their fire went out. In medieval times, a *fire station* might be a special department under central inquisition ready to cremate any witches or heretics. In Babylonian culture, a *fire station* might be a station where all the personnel had fire signs in their horoscopes, and so on.

So, according to Fauconnier & Turner (2003) the overall meaning of composite expressions cannot be predicted from the semantics of the lexical entities comprising them. In other words, their meaning exceeds in principle the sum of their respective parts. Therefore, meaning is not a result of a logical rule of composition but a result of an online integration of minimal linguistic cues provided by the input spaces represented by the components and the frames delivered by the context in which the expression actually occurs. It is a result of a cognitive activity that should involve the semantic features and also the whole phenomenological setting of these features.

Although this approach can be considered a helpful device for modelling crucial types of cognitive processing, it also receives severe criticism, especially in the framework of cognitive semantics. Bundgaard, Stjernfelt & Østergaard (2006) show that the main tenets of the Conceptual Integration theory underestimate the fact that the language is not just a formal system that combines syntactical forms linearly (as traditional compositional theory would have it), but also (and primarily) a symbolic means to express genuine significations (thoughts, cognitive representations, etc.). In Conceptual Integration theory, semantic forms of composition (frames, schemas/scripts, schemata) that serve a semantic binding function are not necessary. However, since standard compositional rules do not seem to work (see the discussion in Section 2.5.2), they claim that the only way to explain why the cognitive processing of compound semantics takes place so easily and automatically, and without being noticed, is that it is strongly guided by such frame- or schema- semantic structure. Following the opinion expressed by Bundgaard, Stjernfelt & Østergaard (2006), I believe that representing this process in terms of conceptual integration without systematically going into semantic details is not enough: the model of conceptual integration ('blending') defines a cognitive processing routine. However, it does not provide a description of either the object being processed (compounds) or the kind of contents expressed by this object. What Conceptual Integration theory allows us to see is the way in which the cognitive system constructs the specific meaning of each of the analysed expressions. It achieves it by postulating that some specific meaning *emerges* or is constituted in the blend by means of conceptual integration. In this approach meaning construal is viewed as a dynamic cognitive process. The linguistic phenomenon itself (the *object* that triggers the cognitive processing in the first place, i.e. compounds in our case) is left out of this analysis.

This means that we only learn that the difference between the expressions *hand wound* and *gun wound* emerges in the blend, although further specifications are not provided. However, if there is

a difference between otherwise parallel constructions, then the lack of an explanation cannot be considered satisfactory and the difference should be addressed in the linguistic analysis.

In the light of the above criticisms, Bundgaard, Stjernfelt & Østergaard (2006) formulate some fundamental rules that in their opinion govern semantic configuration (compound formation). On the one hand, they seem to build on Bauer's (1979) suggestion that all compounds are compositional, because all the linguistic information required for interpretation is present in the compound. Their approach also appears to be somewhat similar to Zimmer's (1971) idea of one-for-all Appropriately Classificatory Relation, which is further developed in Bauer (2006). Bauer's (2006) *Mnemonic Theory* posits that all relations that may occur in nominal compounds can be accounted for with the help of the generic paraphrase 'A type of element-2 efficiently brought to mind by mention of element-1'. The semantic relationship must be "[…] positive, non-modal, and inherent or permanent". This may imply that, for a given compound only one possible reading can occur at once and interpretations that go beyond the limits of the compound's semantics are improbable. For example, the compound *picture book* can be interpreted as a) a book which has pictures; b) *a book without pictures; c) *a book which may contain pictures; d) *a book which contains pictures just today (Bauer 2006: 495-496). The main advantage of this paraphrase is that it is very generic and thus can be applied to all N+N compounds. At the same time it automatically rules out unnatural meanings, and ensures that the reading will be easily perceived by the hearer.

Another important point in Bundgaard, Stjernfelt & Østergaard's (2006) approach is that they also seem to further elaborate Ryder's (1994) idea that in the process of coining a new item, the head concept receives priority, because the head element serves as the referential basis for the expression (see Section 2.6.4.6 for discussion). These rules are summarised below.

(1) Linguistically, compounds are *asymmetric* in the sense that in *XY* compounds, *X* and *Y* do not contribute meaning in the same way. In Germanic *X* serves as some sort of a 'predicate' for *Y*; i.e., *XY* is a construction that prompts the hearer to fit the meaning of *X* into a 'slot' in a schematic frame evoked by *Y*. Therefore, on the semantic level, *X* can always be read as a predicate for *Y*: *X* specifies *Y* in some respect. Inverted compound pairs (e.g. *house-boat* vs. *boat house*, where the former is a boat with some features of a house, while the latter is basically a house having something to do with boats) support this.

(2) Compounds are *constructions* in a sense close to Goldberg's (1995) understanding. Goldberg (1995) views constructions as a kind of correspondences between form and meaning. The form has a meaning independent of what instantiates it. In the case of compounds, whatever concept is

used in the head position, it represents the general focus of attention or the overall semantic frame that is elaborated in the construction. Whatever appears in the position of the modifier should be linked to this frame and should elaborate this frame in some respect.

(3) As Bundgaard, Stjernfelt & Østergaard (2006: 372) claim, this rule is strongly compositional because "it imposes a configurational principle that is invariant through all possible, empirical instantiations of the *XY* construction and fundamentally independent of however *X* and *Y* are construed separately." As the frame evoked by the head is fluid and dynamic (and its meaning may depend on a number of factors including the on-going discourse, contextual meaning and implied information), this relation existing between X and Y does not necessarily have to be simple and predictive.

(4) The way in which the *X* specifies the *Y* depends on two types of theoretical description: *(a*) relative to the general schema instantiated by *X*; (*b*) and relative to the way the integration of *X* in *Y* is cognitively processed in each empirical case.

Some of the ideas expressed by Bundgaard, Stjernfelt & Østergaard (2006) are beneficial for the current research and I will attempt to develop them further. However, it is difficult to agree with them in their complete denial of conceptual integration as part of the meaning construal in nominal compounds, since a number of works on nominal compounding provide empirical evidence that suggests that processing of compounds involves integration of some sort (Bien, Levelt and Baayen 2005, Hampton 1987, Juhasz, Inhoff and Rayner 2005, Libben 2005, Taft 2003, Zwitserlood 1994). Both constituents of endocentric N+N compounds contribute to the meaning in their own way: the head identifies the syntactic and semantic category of the whole and the dependent element specifies the meaning of the head. I also agree with Sandra's (1994) claim that speakers can analyse the whole structure and that the parts are related to the overall meaning of a compound (which does not contradict Bundgaard, Stjernfelt & Østergaard's (2006) claim). However, as discussed in Section 2.6.5, I hold the opinion that endocentric compounds may vary in the degree of their semantic transparency (e.g. *snow shovel* is a kind of a shovel, but *snowman* is not really a man) and compositionality, which I expect to influence the way the meaning of compounds is accessed in the process of online communication.

It is necessary to point out that despite a considerable number of works on conceptual integration, not much is yet known about what it entails. There are two main opinions in the research that discusses the process by which the meaning of a compound is derived. The proponents of the first one claim that the process of meaning retrieval should incorporate decomposition of a compound

(Juhasz, Inhoff and Rayner 2005). The other opinion (Bradley 1980) posits compounds as noncompositional elements of the language, which are represented and accessed as single, indivisible units. Taft (2003: 17), following Langacker (1987a: 461), claims that some part of a compound meaning is derived from the meaning of its constituents, pointing out that some other information is required as well. Unfortunately, little is known about what kind of information this might be, how it could be applied on the level of a compound and how this information is used.

There are two main views on the matter of how the constituent integration process works. The first is that constituent integration involves only activation of the constituent representations and the second is that it is a more interpretative process involving semantic and/or pragmatic analysis as a result of which the compound's meaning is construed.[59] These views will be discussed in the following two subsections where I will look into how the issues of integration are approached in compounding research. I am not going to accept and/or use all of the principles outlined below because I have some reservations which will be discussed as we go. At the end of these two sections I will summarize the ideas that are helpful for my research.

2.6.4.7 *Conjunctive approach*

The main claim of the first of the Cognitive Grammar approaches discussed above is that processing of a compound requires only co-activation of constituent representations (Libben 1998, Taft 2003, Zwitserlood 1994, etc.). The central tenet of this approach is that processing a complex word requires its decomposition into constituents and that lexical and semantic representations of the constituents are involved in compound processing. The results of a number of psycholinguistic studies reported by Gary Libben and his colleagues (e.g. Libben 2005, Libben, Gibson, Yoon, & Sandra 2003; Jarema, Busson, Nikolova, Tsapkini, & Libben 1999) supports the idea that processing of compounds is faster when they have been preceded by one of the compound's constituents. These findings suggest that processing of the compound involves accessing the lexical representation of constituent words and that compound words can be identified via their constituents. Libben (2005: 2777) discusses this suggestion using the example of the paradigm of compounds with the morpheme *bat* used in the position of the modifier. As associated with a particular comic book and movie hero, this morpheme has acquired considerable frequency as an initial compound constituent. *Batman* drives a *batmobile* and a *batboat*, and rides a *batcycle*. He

[59] 'Construal' as a term was adopted in Cognitive Linguistics (and cognitive semantics in particular) from social psychology where it refers to the way in which people perceive, comprehend, and interpret the world around them. In Cognitive Linguistics it refers to the way people structure their experience through language. As explained in Radden and Dirven (2007: 21) "[C]onstruals are operations that help select the appropriate structural possibility among various alternatives" and they are "… strikingly similar to principles of perceptual organisation".

flies a *batplane*, climbs a *batrope*, wears a *batcape*, and works in a *batcave*. All of these compounds are transparent if one posits that the bound compound modifier *bat-*, rather than the free morpheme *bat*, is the one that is employed in compound processing. The model that Libben and Jarema (2006) propose builds on these findings. The model suggests that all possible alternatives for a compound in terms of its constituents are represented in the mind. For example, for any of the compounds of the *bat + N* paradigm, all possible members of the paradigms that contain the constituent *bat*, the full form of the composed and as well as its decomposed form are represented in the model. All of the representations have bi-directional links to each other, thus the full compound form gets activation from both morphological parsing of the constituents and fully structured form representation of the compound in the mind. If we take a relatively semantically transparent compound (e.g. *hand wound*), both of its constituents should be linked with the compound at the lexical and conceptual level. The lexical level of representation includes information about a word's form, whereas the conceptual level deals with the word's meaning. So, talking about our example, the representations of the constituents *hand* and *wound* should become activated on both levels (lexical and conceptual) during the processing of the compound. This leads to increased activation of the lexical representation of the whole compound because the constituents are lexically linked to the complex structure of the whole compound.

The problem with the conjunctive approach, as outlined in Hampton (1987), is that it does not account for new features (emergent features) that appear on the level of a combined concept but are not always present on the level of the constituents. For example, in *pet bird* one of the emergent features is 'lives in a cage', whereas for *pet fish* it is 'lives in an aquarium'. Judging by the given examples one can argue that 'lives in some secluded space' could be part of the meaning of the modifier *pet*, but other cases like *pet name*, *pet topic*, *pet grandchild* or *teacher's pet* do not reveal this feature.[60]

The conjunctive approach seems to be a 'cognitively upgraded' version of the compositional approach described above. According to the conjunctive approach if the novel coinage is not encountered often enough, the parse decays. One should also take into consideration the fact that, though quite productive, compounds have a relatively low frequency of occurrence as compared to simplex words, which means that the creation and re-creation of the concept access node is not possible unless semantic composition takes place. The key question to address here

[60] This criticism is rather arguable, however, since it can be suggested that these new features are not truly 'emergent' features. It may well be the case that the appearance of some new features on the level of a combined concept is connected with the fact that the modifier simply restricts the scope of reference; it confines what subcategory of the large category, e.g. 'bird'/'fish' we're talking about (which is somewhat similar to the way adjectives can be used).

concerns the mechanism that is used for creating the concept node for a compound word and this is closely connected with the issue of what constituent integration involves.

2.6.4.8 *Integration of semantic representations approach*

The second approach to constituent integration posits that the process of meaning construal of a compound is quite active and involves integration of the semantic representations of the compound's constituents. This approach is to a certain degree based on the meta-model proposed by Schreuder and Baayen (1995, 1997),[61] according to which the primary function of morphological processing is to compute meaning, i.e. to access the semantic level by means of processing the lexical form. Therefore, the form and meaning properties of a word should be represented on two separate (but interconnected) levels. The representations are activated via spreading activation and used to determine the meaning of a word. To illustrate this, Schreuder and Baayen (1995) use an example with Dutch words in which the word-form/access representations of *boek* (book), *boeken* (books) and *-en* (the plural suffix) become activated. The word-form activation spreads to the two concept nodes (BOOK and PLURAL). So the representation of *boeken* can be analysed as the union of the two conceptual representations. As long as this is a union, then there is no need for a separate concept node for the plural form *boeken*. Schreuder and Baayen (1995: 13) suggest that the concept nodes need to be added only when "… the meaning of a complex word cannot be obtained as a union of the representational sets of constituent elements". So when it is impossible for the system to figure out what a novel complex unit means on the basis of the constituent concepts, then a new concept node is required. In this case a trace in the memory appears which is stored as a concept node connected to an access representation. However, a new coinage has to be encountered frequently to remain in the lexicon and to be accessed on the basis of its full form, as new representations are usually subject to decay (i.e. to being forgotten).

According to Schreuder and Baayen's (1995) meta-model, family frequency effects are accounted for at the level of conceptual representations. For example, if we take the concept THINK, the concept node for it is activated when we encounter words like *thinker, thinking, unthinkable*, etc.

[61] This model was created for the discussion of inflected and derived words, but its principles can be extended to compounds.

So the more frequently we encounter words containing the same base morpheme {think}, the easier it is for us to access the concept node for the concept THINK.[62]

It seems possible to extend the meta-model's (Schreuder and Baayen 1995) prediction to compounds. Following it, the concept node for OIL should be activated when we encounter compounds containing the morpheme {oil}, e.g. *oil slick, oilskin, oil lamp, lamp oil, sunflower oil*, etc. The question is whether the internal relational structure of compounds can influence the activation of the concept OIL as the meaning of the whole compound is dependent on the relations between the constituents. Even in cases where we have the same meaning of constituents (and, consequently, the same concepts) but they are in the different order (e.g. *oil lamp* and *lamp oil*) the meaning of the whole changes. Taking this into account Gagné and Spalding (2009) suggest that the activation of the concept node is sensitive to the role played by the constituent concept and in this case processing of the compound *doghouse* might be affected by only those family members that have {dog} in the first position (*dog collar, dog biscuit* but not *housedog*). The results of their empirical study show that not only is the position of the constituents important, but also the semantic relation between them. Thus, the meta-model cannot be applied in full to compounds and needs to be adjusted. Response times for targets sharing the same first constituent and the same relations with the primes were notably shorter than when the relations were different, e.g. *snowball* (MADE OF relation) was recognised much faster when preceded by *snow fort* (MADE OF relation) than by *snow shovel* (USED FOR relation). This finding leads the authors to the conclusion that compound processing involves integrating the modifier and the head noun into a relational structure, which presupposes that conjunctive activation of the constituents alone is not enough for the meaning construal of a compound. It should involve a more interpretative process, which includes semantic analysis.

Bien et al. (2005) suggest an intermediate position between the fully decompositional (Levelt, Roelofs and Meyer 1999, Levelt 2001) and nondecompositional (Bybee 1995) models and assume the structured storage view. The main idea of structured storage is that a compound, e.g. *handbag*, might be stored not only as two independent morphemes on the level of form, but together with certain information (including formal, semantic, etc.) about their combination. In this case positional frequency of occurrence of the constituents (the frequency of *hand* as the

[62] It is not clear, however, what the picture will be if we deal with *thought*. Will the concept node for THINK be activated in this case or a concept node for THOUGHT as a noun? How about *thoughtless* and *thoughtful*? A possible suggestion here could be that different concept nodes should be activated since the lexemes are different. But at the same time *thought, thoughtless* and *thoughtful* belong to the same family with *think, thinker* and *thinking*. How are they related? Will the formal expression influence the activation of the base concept?

modifier constituent and *bag* as the head of any compound) is important (Bien et al. 2005: 17877). This seems to be well-justified since language memory blocks are formed by using combinations from prior language experience (which frequency of occurrence reflects on a larger scale). So a word has the ability to be combined with other words that have already been used in the same structural or semantic context.[63] With compounds the structural context is represented by the position of the constituent in a word. For example, the modifier constituent of the compound *handbag* is much more likely to occur in the initial position rather than in the final position. The semantic context is represented by the relation between the constituents realised in a given compound. It is also important to bear in mind that semantic relations even in transparent compounds can be quite ambiguous (e.g. *summer money* – 'money earned during the summer', 'money saved for the summer', 'money lost during the summer'), and sometimes even completely opaque unless the whole situational context is known, especially in cases with novel situational compounds like *Evian boy* with the interpretation 'the boy that helped the speaker open a bottle of Evian water'. And even in simpler, perfectly conventionalised examples like *honeymoon* the meaning is utterly associative, metaphorical and cannot be analysed on the basis of the compositionality principles. At the same time there does not seem to be any contradiction between usual and occasional semantic relations. The possibility of realisation of an occasional relation is connected with a person's ability to produce new expressions and is based on language experience in general, though the combination may not have been encountered or produced before.

2.6.4.9 Summing up integrational approaches to compounding

As shown above, the first of the discussed approaches (a conjunctive activation approach) to constituent integration builds on the assumption that constituents can be used to access the compound's lexical and conceptual representations directly. The meaning of a compound is obtained by means of activating a conceptual representation of the object or phenomenon denoted by a word. According to this approach, the processing of the meaning of a compound involves accessing the lexical representation of constituent words. Compound words can be identified via their constituents and the representations of the constituents should become activated on both lexical and conceptual levels during the processing of a compound. The main problem with this is that it does not explain how new features (emergent features) that appear on

[63] In the case of compounds we can talk about microcontext, i.e. the context provided by the constituents of a complex word. Microcontext is created by the way the constituents of a compound are combined and the relation that appears as a result of this combination. Thus, microcontext is dependent on the semantic content of a word.

the level of a combined concept arise and, as a result, cannot provide a sufficient explanation of how the meaning of a compound word is obtained during compound processing.

The integration of semantic representations approach uses Schreuder and Baayen's (1995, 1997) meta-model, according to which the primary function of morphological processing is to compute meaning, i.e. to access the semantic level by means of processing the lexical form. This means that integration of the constituents should be viewed as an interpretive process, which includes semantic decomposition and analysis of the relational information between the constituents of compounds by an individual.

In the view of the issues interesting for the current research, it is important that when coining a compound, we need both the connection between the lexical and conceptual form of the constituents of compounds as well as understanding of the expected output. For this we need to be able to introduce relational information that would allow a compound to acquire a certain meaning. In order to be able to do that, it is necessary to assume an intermediate position between the fully decompositional (Levelt, Roelofs and Meyer 1999, Levelt 2001) and nondecompositional (Bybee 1995) models. This raises the question of whether the information about the semantic relations is pre-stored in the concept or whether semantic relations form a group of a certain type of relational concepts that emerge in a structure when a need arises. This question is important for a current research and will be discussed in the following section.

2.6.5 Reconsideration of semantic relations

Estes and Jones (2006) provide some interesting insights into the issue of relations between concepts in N+N sequences. They are mainly concerned with the question whether such relations are represented as independent units in the semantic network (model of independent representation), or whether they are represented as part of the particular concepts that instantiate them (model of bound representation). In the model of bound representation, a relation is represented as part of the meaning of whatever concepts entail that relation (Gagné 2001, Raffray, Pickering, Branigan 2007). For example, the MADE OF relation of COPPER HORSE is represented as part of the meaning of COPPER. Gagné (2001: 247) argues that ''relations are associated with the modifier's representation, rather than existing as independent structures''.

An alternative position is offered in the model of independent representation, which claims that relations are representational structures themselves (Estes 2003, Spellman, Holyoak, & Morrison 2001). According to this model, the MADE OF relation, for example, is represented independently

and is not part of any particular concept, e.g. COPPER, GOLD, CHOCOLATE or any other concept. This relation may be activated by those concepts, but it is not necessarily a part of their representation, but is an independent representation in and of itself.

The underlying idea for the independent semantic relations as separate concepts is based on the fact that combined concepts are understood by retrieving or inferring some relation between the given concepts. A number of experimental studies have shown that understanding of a stimulus (e.g. *animal cage*) can be facilitated by a prime which has the same relation as the target (e.g. *fish pond*). If relational representations were concept bound, then the relational priming effect should only occur when the prime and the target are somewhat similar lexically. This, in fact, happens as well. In a study by Gagné (2001) one can see that the lexical priming effect is especially strong when the prime and target have the same modifier (e.g. *bird nest* → *bird cage*). Nevertheless, relational priming occurs in the absence of lexical repetition (Estes 2003, Estes and Jones 2006, Raffray, Pickering and Branigan 2007), as long as the preceding combination utilises the same semantic relation as the following one (e.g. *fish pond* → *bird nest*). Moreover, as pointed out by Estes and Jones (2006: 90), the models (of bound representation and of independent representation) differ dramatically in their representational demands. For example, *bear paw* and *fan blade*, both instantiate the POSSESSION relation. If we accept the position of the bound representation theory, then the relations should be bound to their particular concepts of instantiation, which suggests that the elements (e.g. *bear*, *fan*) must represent this relation separately as part of their individual conceptual structure. In fact, this relation must be represented in every concrete concept[64] separately. Developing their argument, Estes and Jones (2006: 90) also note that we should expect just about every concept to include the CAUSE relation as part of their representation, since just about any concept can be involved in a causal relation (e.g., LOTTERY RETIREMENT). In other words, the representation of every concept in one's semantic network must include every relation that the concept could possibly instantiate, which means that the model of bound representation is unnecessarily redundant. Estes and Jones (2006) demonstrate this redundancy analysing the paradigm of *bear + N* compounds, in which a variety of relations is instantiated: *bear paw* (POSSESSION), *bear scare* (CAUSE), *bear season* (TIME), *bear toy* (POSSESSION or ESSIVE), *bear tracks* (SOURCE), *bear cave* (PURPOSE), *bear story* (TOPIC), etc. If we assume that every single concept potentially may appear in a combination where any of these relations can be instantiated, then the same set of relational concepts should be represented in the conceptual structure of every noun, which means that this information should also appear as

[64] Since we generally expect concepts that relate to concrete objects to consist of one or more parts.

part of the meaning and should be reflected in the semantic content of every noun concept. This means that such redundancy would be extremely taxing in terms of representational demand.[65] This, presumably, would also put strain on processing, since the time and effort required for searching through those representations increases with the number of concepts.

The discussion of the empirical evidence that supports the suggestion of the model of independent representations of semantic relations is ongoing. The results of experimental studies reported in a number of papers (e.g. Estes 2003, Estes & Jones 2006, Raffray et al. 2007, Spellman et al. 2001) present evidence that relational priming occurs in the absence of lexical repetition, i.e. the compound *bird nest* can facilitate processing of the compound *turtle cage* because both combinations utilize the same relational representation.

However, the 'bound' and the 'independent' models do not necessarily have to be mutually exclusive. It might be that we can have knowledge of kinds of relations on the one hand AND have associations attached to 'things' (nouns) as to what relations they typically engage in.

In the 'independent' model relations are viewed as independent entities, each relation needs to be represented only once. Estes and Jones (2008) also use the model suggested in Leech, Mareschal and Cooper (2008) to explain why semantic relations should be pre-stored separately in the lexicon and should be used in conceptual combinations on the principles of analogy. Leech et al. (2008: 363) claim that analogies are understood by relational priming, and according to them, many cases of analogy involve seeing the similarity between the prominent relation in one domain and the prominent relation in a second domain. For example, exposure to related terms, e.g. *puppy* and *dog* should activate a relation between the terms (e.g. offspring), which may then biase the term *kitten* to produce the corresponding term *cat*. Leech et al. (2008) also posit that relational priming also occurs regularly in common language use, of which nominal compounds are a good example. As discussed in Section 2.4, understanding of a nominal compound involves inferring a semantic relation that holds between the constituents. A number of studies report the relational priming effect in understanding conceptual combinations by individuals. Relational priming occurs when one phrase (e.g. *bird nest*) facilitates comprehension of a sequence that

[65] So assuming a model of economical storage in memory seems somewhat preferable. It can be explained by using an analogy with the dictionary, where definitions contain only necessary and sufficient information. This means that if the mental lexicon (a metaphor in its own right) were a list of entries like a dictionary, then it would definitely be more economical to ban 'relational information' from the nouns, simply because it would make the entry per noun longer than necessary. Listing just a limited number of relations as separate entries instead appears to be a more economical option. However, it is still to be proved that this may be considered a realistic metaphor of how the brain stores, retrieves and processes language.

utilises a similar relational pattern (e.g. *fish pond*) (Estes 2003, Estes & Jones 2006, Gagné 2001, Spellman et al. 2001). Leech et al. (2008) claim that the connections between analogy and conceptual combination allow for some useful comparisons between the research in the two areas.

Leech et al. (2008) claim that within the context of analogy, relations may be conceptualised as a certain type of transformations between items. It seems that Leech et al.'s (2008) use of the term may be somewhat misleading and requires further clarification. By transformations Leech et al. (2008) mean that the relation is represented as the pattern of activation required to transform an input object (e.g. *nest*) into an output object, which is a compound concept (e.g. *bird nest*). In this case transformations are understood literally as ways of changing/specifying/elaborating the meaning of the head concept of a compound by means of applying a semantic relation to the coined structure. So the meanings of the constituents are modified with the relations. As noted in Estes and Jones (2008), this transformational model of relational representation explains some of the observations in research on conceptual combination. One of such observations is that processing of both familiar combinations (e.g. *bird nest*) and novel combinations (e.g. *turtle cage*) appear to utilise the same processes. It is a generally accepted that familiar combinations would be understood by simply retrieving the compound concept from memory, whereas novel combinations require processing of meaning and inferring a semantic relationship. However, some of the recent studies (e.g. Gagné & Spalding 2004) suggest that both familiar and novel combinations undergo the same processing operations. Leech et al.'s (2008) transformational approach explains it in the following way. The transformation from simple concepts to a compound concept requires inferring a semantic relation, and the familiarity of the compound is important only in the sense that the transformation may be faster for familiar compounds than for novel compounds. Leech et al. (2008) note that inferring the relational information still must occur in both cases. This means that the process of coining a novel conceptual combination may somewhat be related to the process of producing a familiar item in the online communication. Both processes require understanding in what way the combined concepts are connected but this process will proceed more quickly with more familiar items. The idea that relational representations may be independent of the concepts that instantiate them may be helpful on this respect as well, as it allows for viewing semantic relations as a certain type of a relational kit, from which a required relation can be picked. The advantage of this then is that relations do not have to be presented explicitly, thus the difficulty of learning explicit structured representations is avoided.

Semantic relations as conceptualised units, pre-stored separately, should bear the traces of meaning typical of other concepts they have ever been used with (in the same way concepts should bear the traces of relations they have ever been associated with). As long as the usage of semantic relations is driven by analogy, the traces of meaning from other concepts have to be stored in memory as associations. These 'leftovers' can be used for creating the meaning of a compound if applicable. For example, the MADE OF relation can be found in compounds like *glass vase, metal bar, stone wall, plastic bag*, etc. The first component in compounds with the MADE OF relation denotes a substance from which the head component is made. So, when we encounter a novel compound with the modifier naming some sort of substance, the MADE OF relation is probably most likely to be applied and the MADE OF relation, in its turn, should be associated with concepts representing substances. At least it is expected to be the case for semantically transparent compounds.

The model of independent representation of semantic relations looks quite promising and logical as in this case the information stored in the lexicon is compact, which provides easier access to it as compared to the models that presuppose that semantic relations are directly connected and are parts of other concepts. The idea about semantic valences being the relational potential of noun concepts seems beneficial, since in this case the semantic content of the concept is not overloaded with the information that is irrelevant for a concrete situation. In other words, if we presume that all nouns have the same valences, there is no need to list those valences as discriminatory information in a dictionary-like mental lexicon. The knowledge that nouns have those valences is probably stored together with the other general information that we have about nouns both grammatical and semantic, i.e. that they denote things, that they can be singular or plural, etc. It also correlates with the idea that semantic relations can be considered separate entities that are not included in the semantics of a lexical unit. The reasons for this are suggested by Estes and Jones (2006), whose priming experiments demonstrated facilitation effects in the absence of lexical similarity. A study by Bauer and Tarasova (to appear) demonstrates that the same set of semantic primitives seems to be recurrent in a number of nominal constructions. This allows for a speculation that semantic relations that exist between units of a language are, in fact, independent conceptual entities. This approach of understanding semantic relations as separate entities should imply that they have separate entries in the lexicon. The use of these entries is required in cases when decoding of semantic information packed in a construction involves analysis and decomposition. The set of relational concepts is assumed to be fixed and finite (but can be extended in the course of the language development through time).

If we assume the position that semantic relations are not bound to the concepts that instantiate them, but are rather independent entities, then it is not clear how and why this relational information is inferred into the combination of two adjacent nouns, i.e. what it is in the noun concepts that allows for this inference as well as for a variety of semantic relations that can be used in N+N conceptual combinations. On the one hand, we know that the main purpose of noun concepts in general is to designate entities, but not how entities are related to each other (as is the case with adjectives, for example).[66] On the other hand, N_1 in an N+N sequence functions as an attribute, which is not a very typical function for a noun (but is very typical of an adjective). It seems that this clash between the purpose of the noun concept in the language and its actual functioning in the attributive role may be responsible for the possibility to infer a semantic relation between the constituents in order to produce the intended meaning. However, it would be interesting to see if the possibility of using a noun as an attribute in N+N sequences makes it in any way similar to adjectives that are used in their natural role of an attribute. In the next section I will look at the possibility of drawing parallels between Adj + N and N+N sequences in order to understand where semantic relations come from.

2.6.6 Adjective + Noun sequences as compared to N+N compounds

The results of a study by Medin and Shoben (1988) on conceptual knowledge associated with Adjective + Noun (Adj+N) conceptual combinations suggest that the combined concepts reveal semantic features that are not typical of or central to the concepts taken separately. For example, it was discovered that the participants of their experiments were more likely to associate *metal spoons* with the concept *small spoon*, but *wooden spoons* were usually associated with that of *large spoon*. At the same time, spoons as such are typically judged as small.[67] In the conjunctive approach to constituent integration (i.e. simple constituent intersection), the conceptual representations corresponding to the combinations *wooden spoons* and *metal spoons* should include only sets of attributes of the constituent concepts, and in this case the representations should be free from size dimensions (LARGE and SMALL), which is not the case. As long as we can see that size dimensions are still present in the mental representation of the new concepts, it does not seem right to talk about pure constituent intersection. The concept of frame could be helpful in understanding why combined concepts reveal semantic features that are not typical or central to the concepts when they are taken separately. To speakers with the appropriate frame, it is not

[66] It is worth pointing out that entities denoted by relational nouns like *mother*, *brother*, etc. cannot be conceived of without understanding of the relation in which they are related to other entities.
[67] This might be considered to be evidence for the prototypical spoon being small and probably metal.

surprising that wooden spoons are perceived as large spoons. For example, to speakers in North America, Western Europe and elsewhere wooden spoons are used for cooking; therefore, they have to be large so that cooks can stir food in pots without getting burned. Metal spoons are mainly used for eating.[68,69]

The results of Medin and Shoben's (1988) study can be extended to N+N compounds. Like N+N sequences, Adj+N structures may transform into compounds (e.g. *blueberry*, *blackboard*). Adjectives possess a number of prototypical features. From the viewpoint of syntax they are used as premodifiers (e.g. a *black car*, *small room*) or predicates (e.g. *this car is black*, *that room is small*) and are heads of adjective phrases. Semantically English adjectives are claimed to be ascriptive; that means that they denote 'a property which is valid for the entity instantiated by the noun' (Ferris 1993: 24), i.e. in the combination *beautiful picture*, *beautiful* is the property of the picture. Typical adjectives are also intersective (Siegel 1980), i.e. *beautiful picture* is a member of the intersection of the set of pictures and the set of beautiful objects. Quite a few adjectives do not demonstrate this typical behaviour. This does not seem to represent any problem with the morphological status of adjectives denoting dimension, which are not intersective (*small*), adjectives that are not gradable (*dead*), etc.

Adjectives are not homogenous and can be divided into many various types and subtypes. This is also true for nouns. In many cases, e.g. *chocolate, lemon, future, animal*, it is difficult to say what we are dealing with: a noun or an adjective, since dictionaries list them as both. So, not only is there no single sharp line that we could draw to distinguish between the two classes of words, it is hardly possible to draw the lines that would go between different kinds of nouns and adjectives.

It can be argued that Adj+N combinations provide an easier access to an individual's memory. According to Langacker (1987b), adjectives can be referred to as relational entities since they are linguistic expressions that designate relations on a par with verbs, adverbs and prepositions; unlike nouns the main purpose of which (apart from relational nouns like *mother*, *father*, *sister*,

[68] This frame will probably not be the same to speakers in Russia and other countries where wooden spoons may still be used for eating as well as for cooking.

[69] The difference in the content of prototypical concepts (SPOON in our case) and the content of combined concepts (e.g. WOODEN SPOON and METAL SPOON) is a matter of heated arguments in Cognitive Linguistics literature, and is part of an argument whether prototypes (as a theory of concepts) should be rejected altogether, e.g. Allan 2009, Connoly, Fodor, Gleitman and Gleitman 2007, Cruse 1986, Croft and Cruse 2004; Fodor 1998; Fodor and Lepor 1996, Taylor 2008, Jackendoff 2002, etc. One of the most common criticisms in this respect is that prototypes cannot account for meaning composition in conceptual combinations in general (and N+N compounds in particular). For example, *pet fish* cannot be a combination of the prototype of PET and FISH because a goldfish, which is a typical pet fish, is neither a prototypical PET (furry, cuddly etc.) nor a prototypical FISH (e.g. herring). The prototype theory is argued not to provide a sufficient explanation as to how one can gain access to the meaning of a compound through its component parts if the elements of meaning present in the compound are not present in the components.

etc.) is to designate entities.[70] The term a 'relational entity' suggests that there should also exist a nonrelational entity. Cruse and Croft (2004) provide the following explanation. On the example of adjectives and nouns we can see that an adjectival concept SMALL cannot be conceived of without a reference to what is being described as 'small', thus it is relational. Nouns, whose function is to designate nominal concepts, can be conceived of as such, without a reference to another entity, i.e. a nominal concept TABLE can be conceived of without there being any reference to some other entity. Langacker (1987a: 198) suggests that nouns ('things' in his terminology), or, rather, nounhood, construes a concept as a region or 'set of interconnected entities' and entities are nonrelational. Relational concepts (in our case adjectives) profile the interconnections between entities; nonrelational concepts (nouns) profile the entities that are interconnected (Langacker 1987a: 216). Things are autonomous and are relatively stable in space and time and the same can be said about nouns that name these things.

Describing nouns as labels for construed entities seems to be a rather accurate definition of nounhood and should to a certain extent resolve the issue of the morphological status of the modifier in sequences like *table tennis* or *door knob*. It can be questioned, though, whether *stone* in the combination *stone wall* is a concept denoting an entity or a concept denoting a relation, as it is usually defined in dictionaries as both a noun and an adjective. Besides, how is *stone* in *stone wall* different from *wooden* in *wooden floor*? But taking the cognitive approach we can see that there is no problem for us to conceive *stone* taken separately as a material entity, whereas *wooden* requires something to be described as wooden.

It is necessary to note, however, that one of the questions that can be asked here is that attributive adjectives, in a noun phrase, do require a noun. However, in Langacker's approach it is not clear whether the necessity for a noun to be present in a noun phrase is a syntactic property of adjectives or whether it is a conceptual one. That is, in the case *A stone was lying in the road*, we deal with a complete piece of syntax, that does not pose any difficulties in terms of its interpretation. Examples like *A stone with a huge mane guarded the gates* might be syntactically complete but not so easy to interpret. In the latter example, the context provided by the sentence makes it possible to suggest that the word *lion* might be missing. The situation may become more complex to account for if we think of examples of substantivised adjectives, e.g. *woollens*, *delicates*, *sharps*. However, depending on the assumed position, the use of adjectives to replace a

[70] The use of the term "relational" may be confusing and it is necessary to point out that Langacker uses the term in the sense different from that of Levi (1978). The notion of relationality is understood by him as a tool for distinguishing between entities that cannot be understood without being related to some other concept (i.e. adjectives and verbs) and entities that can be understood without the necessity to be related to another concept.

noun phrase may also be considered a case of metonymy, where (some quality of) a material noun replaces the entity made from this material. This means that the analysis should shift its angle, since we deal with figures of speech (some of which may even become conventionalised in the language, as is the case with cases like *woollens*, *delicates*, *sharps*, etc.).

So, the difference in the conceptual structures of nouns and adjectives can be connected with adjectives being relational entities (Langacker 1987a). Since designating relations is the main function of adjectives, the information about the type of the relation is included in their semantic content. Therefore, when encountering an adjective we relate it to something it can modify (most commonly nouns). THINGS can be conceived as such, independently of a relation. Nouns are the prototypical means to refer to what is construed as a THING. When nouns modify other nouns (as in the case with N+N compounds), though they are functioning in the way adjectives do, the relations that are activated are not obvious, as they are not so limited in combinability with other nouns.

2.7 Outline of the chosen framework

The following sections look at the questions important for explaining how compounds are coined. I start the outline of the chosen framework with the proposal about how it is possible for a sequence of two nouns to acquire a relational reading and what can be responsible for that. I reconsider the nature of semantic relations and offer a different outlook, which in my opinion, allows for a more accurate explanation of how and why a relational reading is inferred. The offered approach uses the semantic content of the constituents as the basis for the prediction a relational meaning that may possibly be inferred into an N+N sequence. As is pointed out in the beginning of this thesis, there is hardly any research done into the issues of either coining of compounds or their online processing. Therefore, all suggestions put forward in this thesis utilize the knowledge obtained in the course of studies of processing of meaning of nominal compounds. It has been suggested in some studies (e.g. Levelt, Roelofs and Meyer 1999) that comprehension of nominal compounds (as well as any other linguistic items) is closely connected with their production. Production could not occur if comprehension was not involved because a novel sequence is produced on the assumption that it is comprehensible for the listener/reader.

This is followed by a discussion of a more descriptive question; that is the question of the productivity of a noun as a constituent of a compound. The discussion of the surface matters is extended to the semantic level and the possibility of a constituent to favour one semantic relation over the others is looked into. Some of the principles of analogical word-formation (that are used in Construction Grammar) are utilized in order to provide a grounded discussion of the productivity patterns in compound formation.

2.7.1 Relational potential of a noun and semantic valences[71]

The discussion in the previous section demonstrated the parallels between Adj+N and N+N sequences, suggesting that nouns have a relational potential. Based on this assumption, it is possible to further suggest that if used in the role typical of adjectives (attribute in our case) nouns may act in a similar way to adjectives. However, acting similarly to adjectives does not mean that they are ascribed certain relational information (which is part of the meaning of adjectives).

However, even if we assume that nouns can manifest some traits of behaviour typical of adjectives, we still need to look more closely at the question of the origin of the semantic

[71] In this thesis I assume the position that is different from that of Taylor's (2002) understanding of valence relations. He talks about valences in syntactic terms and distinguishes three types of valences: head−complement, head−modifier or the appositional relation (see Section 2.6.4.3 for more details).

relations. It is suggested here, that the relational potential of a noun is responsible for the possibility of inferring a semantic relation in order to understand what a given combination of two nouns may mean. The notion of relational potential is strongly connected with the notion of semantic valences, which was introduced and discussed in Sections 2.7.1.1 – 2.7.1.3.

2.7.1.1 *Valence vs. semantic relation*

It is assumed here that, like adjectives, nouns can possess valences, which can be described as the ability of a word to be combined with another word or other words. There is some ambiguity in the literature concerning the terms 'valence' and 'relation' and authors often use them interchangeably. This ambiguity takes its roots from the necessity to differentiate between syntactic valence and semantic valence. The difference between the two can be described as follows. The term 'valence' (first suggested by Lucien Tesnière in 1938) and a competing term 'argument structure' are applied mainly to the description of the surface syntactic relations of a verb. Later the term 'valence' was adopted in semantics and was extended to other parts of speech, e.g. adjectives (*angry* is bivalent: who is angry with whom?) and nouns (whose *sister*? – Herman's). The syntactic relations a word can have are prescribed by its semantics: a word that possesses some syntactic valence has to always be referred to a situation which, in its turn, always has to have participants. These participants are often presented explicitly and fill in the syntactic valences of a given lexical item. As a result, syntactic valences explicate the semantic relations between the situation and its participants.

Unlike syntactic valences, semantic ones are distinguished on the basis of the sense relations they express, which makes them analogous to semantic roles. It is generally accepted that there are about ten semantic roles; however, Apresyan (1974) names twenty-five types of semantic valences. Among these are valences of the subject (*a train moves*), of the counteragent (*protect somebody against something*), of the recipient (*give me*), of the addressee (*inform somebody*), etc. Apresyan's classification can be made even more detailed and the subdivision into types of valences can be limited by the following condition. Valences that are close in meaning can be considered different if they can occur in the valence structure of one and the same word (e.g. the valences of counteragent and mediator: *buy from a firm via an agent*). Syntactic valences are connected with the situation, which requires a verb as a core condition, but semantic ones are not

so dependent on the verbal component and can be considered a property of any notional part of speech.[72]

Nouns can possess valences as long as they are related to other nouns. Taken separately they have the potential for taking different semantic roles (i.e. the most typical semantic relation for the nouns denoting material will be the one of content). But nouns are relatively unlimited in how many valences and of what type they can have due to the creativity of human thinking and the variety of situations where nouns can be used. The restriction here is that nouns, being 'entity words', are relational only potentially. To activate this ability they have to be in a 'relational position', i.e. function as relational entities. In the English language nouns cannot fulfil the same function as verbs or adverbs do, but can still serve as modifiers of other nouns. Dependent components of N+N compounds can be considered an example of nouns activating their relational potential.

I also take the position that valences should not be confused with semantic features, e.g. [+ANIMATE]. The difference between the two notions is that semantic valences are components potentially present in the semantic content of a noun, while semantic features are binary. Valences are open slots that may be active or not so active. The activity of a valence is connected with the semantic features but does not have to be directly determined by it. For example, one of the features the noun *animal* is characterised by is [+ANIMATE]. This feature allows for the connection with the information we possess about animate objects (that they live somewhere (LOCATION), are used for something (INSTRUMENT), etc.), which allows for the activation of the corresponding valences.

2.7.1.2 *Semantic relations as instantiations of valences*

The present study assumes that since there are syntagmatic relations within composite structures that include nouns, then nouns should possess an argument structure, just as verbs do (Plungyan and Rakhilina 1998). However, the argument structure of nouns should be somewhat different from that of verbs, as nouns generally do not project syntactic structure in the same way as verbs. The semantic structure of a verb surfaces in the syntax of constructions it occurs in. Such characteristics as aspect, transitivity-intransitivity, ability to be used in certain grammatical constructions, for example, which are important for creating the syntax of an expression, are

[72] I do not include any introduction to the theory of thematic roles in this thesis, since although very interesting, I feel that the discussion of the predicate argument structure of clauses in the analysis of nominal compounds can lead me away from the points that are believed to be important for the current investigation.

largely determined by the semantics of a verb. This does not seem to be the same for nominal combinations, since the nature of the relations between the components of nominal constructions is more complex (in terms of meaning construal) than the relations between the components of verbal ones (Ryder 1994: 28). This complexity arises from explicit predication that is present in the semantics of verbs but which is not present in the semantic structure of nouns. However, it appears that in the case of nouns, the ability for predication is not altogether absent, but is stored on the level of conceptual representation. This ability surfaces in the interpretational patterns of nominal constructions, e.g. *a red apple* – an apple **is** red, *the boy's hat* – the boy **has** a hat, *a police officer* – an officer that **serves** in the police, etc.

At the same time the lack of overt relations that are strictly prescribed by the semantic characteristics of a noun allow for much more freedom in the process of combination (as well as interpretation). It is suggested here that a noun concept theoretically has an unlimited semantic potential for being combined with other nouns, but this potential does not necessarily have to be realised in the language; and, in fact, it is not, if there is no pragmatic need for such a realisation, i.e. if there is no suitable referent for it. For example, there is no such a compound as rose police because at the current moment of time we do not seem to have an entity that can be denoted by this sequence.[73]

The idea about the unlimited potential allows for an explanation of the possibility of coining novel creative formations or novel readings for mundane constructions (e.g. *firemen* in Ray Bradbury's sense, i.e. firemen set something on fire). The number of unfamiliar combinations, creative combinations (as in Benczes 2006a) or creative uses of mundane combinations is not as large as the number of compounds which are used in their direct meaning.[74]

As pointed out before, the potential may be unlimited; however, only a certain number of valences seems to be instantiated (and the semantic relations corresponding to these valences are

[73] Google search on 20.03.2013 resulted in no hits for *rose police*, which does not mean that the situation cannot change. In the course of revising this thesis the example of *rose police* replaced the older example of *rose mouse*, which had no hits in 2011 but now seems to be used to refer to a cute picture of a mouse holding a rose (as used on postcards, etc.). So, if somebody decides that there should be a police department dealing with criminals that destroys roses in parks, we may have a compound for this and *rose police* could be one of the possible candidates.

[74] It is worth noting that understanding of creativity in compounding is different for different authors. In this research creative compounds are those that utilise metaphor, metonymy and creative associations between concepts (Benczes 2006a: 7). However, it is also important to understand that using a language is a creative process as it is and is virtually impossible to draw a strict borderline between expressions that should be considered creative and the ones that should be viewed as mundane. For example, the use of metonymy, e.g. *wind band*, *business community*, etc. (as discussed in Sections 4.1.7, 5.4.1.5 of this thesis) appears to be a regular process in compound formation and such compounds are not perceived as being creative and/or unusual.

inferred) consistently and repeatedly across N+N conceptual combinations (hence all the attempts to classify the semantic relations in taxonomies). For a linguist the degree of specification of the classification of semantic relations depends on the framework chosen for the description.

The possibility of putting two nouns together to form a compound depends on the semantic content and structure of the constituents, which allows for certain valences that nouns possess to be activated in the process of combining two nouns for creating a new sense. When the semantic structures of two nouns have matching properties (in this case, valences), the possibility of combining them in order to produce an N+N expression should be quite high. However, the nature of 'matching' is not that of sameness, but can be compared to a jigsaw puzzle in which a slot in the semantic structure of one noun concept can be interlocked with a knob that is more or less likely to fit into the given slot. For example, in the paradigm of *N + animal* compounds, based on the semantic content of the head constituent *animal*, one of the valences that can be potentially activated is that of LOCATION, since an animal has to live/be found somewhere. This valence is instantiated as a semantic relation in such compounds as *sea animal, zoo animal, forest animal*, etc., in which the semantics of the modifier constituent determines the activity of the corresponding (LOCATED) valence. So, a better way to view this may be to envisage the noun concept as an uneven sphere, whose protuberances and recesses clip on to other noun concepts. The more often a noun concept is used in a certain combination or series of combinations when one of the valences is activated, the stronger the possibility of using a particular semantic link becomes. As a result of this (based on the data collected for the current research), for instance, in the case with *N + area* compounds, we are more likely to expect the locative relation (N_1 located in N_2) to connect the constituents, e.g. *city area*, *mountain area*, rather than the copulative relation (N_2 is N_1), e.g. *wilderness area*, *slum area*; and the probability of the instrumental relation (N_2 uses N_1), e.g. *technology area*, is even lower.

So, when the concepts are combined, in the process of matching corresponding slots, a relation that keeps the concepts 'glued' together appears. So, in this thesis semantic relations are viewed as results of this process of combination and are somewhat different from valences. Valences are believed to be potentially present in the conceptual structure of a noun and relations are the actual instantiations resulting from the activation of valences in individual combinations. It is necessary to point out that this idea of matching corresponding slots is not new and is based on Ryder's (1994) proposal that the schemas of the constituents should be common or identical in order to be combined into a meaningful unit. Ryder (1994: 72) further elaborates this thought,

suggesting that in the course of communication, language users look for similarities in the schemas of both components in order to find matching slots. This results in the alteration of the meaning of both constituents when a composite structure is created. The process of accommodating the schemas into a new construction should take place when compounds are created and when they are interpreted.

2.7.1.3 *Valences and predictability of meaning of compounds*

One of the 'hot' issues in compounding research is the issue of predictability of meaning of a given compound. The question of predictability is often discussed together with the question of compositionality of compounds (see Section 2.5.2).[75]

Exploring the question of predictability of meaning of a compound, Štekauer (2005) claims that predictability is the degree of probability of how the meaning of a novel language unit will be understood by a speaker/hearer/reader as compared to other possible meanings. Here the context in which the combination occurs is of vital importance. However, verbal context cannot always be enough to provide the accuracy of interpretation and situational context should be enlisted as well (Kavka 2009: 24). For example, the combination *blanket man* will most probably allow the interpretation of 'a man who sells blankets' for someone who has never been to Wellington (or hasn't seen the article in Wikipedia) and only the situational context (which includes cultural background and knowledge of the world by communicants) will provide the necessary basis for a plausible recognition of what is meant by this naming unit, i.e. a homeless man in Wellington who wore a blanket at all times. The other examples could include the ridiculousness of the referent named by the compound *milkman* for Vietnamese learners of English and the specificity of the phenomenon denoted by the compound *sandwich man* for Western countries. This kind of knowledge is important for understanding whether the encountered language unit is used literally or figuratively. This knowledge is necessary for further processing of the meaning.

The current research employs the idea of valence structure of a noun in an attempt to investigate further the question of predictability of meaning of a compound. As is pointed out above, every noun is assumed to have a set of semantic valences that are consistently instantiated as semantic relations. The number of valences is potentially unlimited but not all of them have equal

[75] This point is argued by Kavka (2009: 24), who justly points out that it is not always the case and should not be expected, and these two qualities are not necessarily interdependent. He gives the example of *N+safe* compounds (e.g. *dolphin-safe* vs. *child-safe* vs. *shark-safe*), which can demonstrate unpredictability of meaning.

probability of being activated; in fact, only a few are activated consistently and repeatedly. In the case of compounds, the probability of a valence for a particular noun should be determined by its semantics and the position it takes in an N+N sequence. For example, (based on the data collected for this research) it is rather unlikely for the noun *product* in the meaning 'something produced by effort, or some mechanical or industrial process' (Collins online 2012) when used in the head position to activate the valence that is instantiated as the POSSESSION relation (but still not impossible, e.g. *quality product, value product*); whereas the probability of instantiation of the COMPOSITION (e.g. *milk product, wood products*) or PURPOSE (e.g. *hair product, skincare product*) relations is very high, etc. At the same time, when used as a modifier, the situation is very different, with the POSSESSION relation being the most prominent in the *product + N* paradigm (e.g. *product design, product life, product name*).[76] This brings us to the discussion of two suggestions. First, it may be possible to predict the semantic relation that will be realised in a given compound based on linguistic and extralinguistic knowledge about the constituent. Second, each noun has a set of valences, which is the same for all nouns, but the probability of those valences differs from one noun to another.

The probability of valences and their probability rate (henceforth – PR) is not strictly prescribed but is constantly changing with the development of the language and appearance of novel constructions. It is virtually impossible to stop speakers of a language from using a word in a sense that was not typical of it before and thus raising the 'rank' of a given valence. For example, the word *potato* started to be used in the figurative sense to refer to a person who spends too much time doing something that does not require intensive physical activity with the appearance of the term *couch potato*. The latter gave grounds for analogical formations like *mouse potato* and *web potato*. Thus, we can say that because of developing a new sense, the valence allowing for the formation of these combinations is more active now than it was before.

It may be helpful to represent the probability rate of valences (PR) of a noun on a scale from 1 to 3, with 1 being the valence that is very predictable and 3 being the valence that has the lowest rate of predictability. The differences in the PR are determined and defined by the structure of the concept, which is reflected in the semantic structure of a noun representing this concept in the language. A number of other factors (mainly based on our pragmatic knowledge of the subject) are also important. It may be possible that the PR for a given noun differs from one individual to

[76] The directionality of the POSSESSION relation is also different in these two paradigms (N_2 HAS N_1 in *quality product, value product* vs. N_1 HAS N_2 in *product design, product name*, which is another interesting issue that I hope to look into in a separate investigation.

another, since one and the same concept may have different value for different people. However, for the sake of argument, we will not be concentrating on factors such as the importance of a concept for an individual, the emotional experiences associated with the concept, etc. in this research. The PR is not once and forever established; it may change depending on which aspect of a concept becomes more important for the speech community.

2.8 Productivity patterns in N+N compounds

Another issue that is of interest for current research is connected with a claim of a more descriptive nature, i.e. that compounding, being mostly a morphological process, is productive and some elements are more productive than the others. Bearing this in mind, Hamburger and Jacobsen (1972) suggest that the constituents comprising a compound in Danish should be more productive either in the position of the head or of the modifier. This seems to be the case for English as well. For example, the word *water* used as a modifier occurs much more often and in a wider variety of combinations than as a head noun. In other words, it has a larger modifier family size. According to the data extracted from the BNC, the modifier family for the constituent *water* contains 48 types (e.g. *water authority, water industry, water services, water board, water charges, water temperature, water condition, water garden, water shares, water park*, etc.), whereas the head family contains only 15 types (*sea water, surface water, bath water, salt water, waste water, rain water*, etc.). In the opposite case, the word *problem* can be found in the head position in 116 of compounds (*back problem, crime problem, money problem, business problem, quality problem, family problem, disease problem, city problem attitude problem, acquisition problem, housing problem,* etc.), but its occurrence as a modifier is very limited, with only 10 compounds where it is used in this role (e.g. *problem behaviour, problem children, problem situation, problem area, problem families*, etc.); i.e. its head family size is much larger than those of the modifier. The question of productivity of constituents has not been widely discussed in the literature and most works concentrate on productivity of compounding as a process. However, looking at the issue of productivity from the position of the constituents comprising a compound may help shed light on the question of how compounds are formed.

2.8.1 Structural analogy[77] and its influence on the productivity patterns

One of the questions currently discussed in the literature in relation to analogy in word-formation is how to distinguish between rule-governed and analogy-based formations. In a number of modern approaches analogy is viewed as the only morphological mechanism available.[78] Such approaches define analogy in terms of the kinds of features over which similarity is systematically computed over sets of words. This process of similarity computation is implemented in computational algorithms and can be used to predict, generate or select pertinent complex words

[77] By structural analogy I mean word order on the level of an N+N sequence.
[78] See Bauer, Lieber and Plag (to appear), Chapter 29, for the discussion of the analogical models.

as the output of a morphological category. A considerable number of studies[79] have provided very promising results, with some of them clearly indicating the superiority of analogical models to competing non-analogical models.

One of the opinions that are widely argued in the literature is whether rules and analogy are fundamentally different in their nature and mode of operation because they are supported by different mental mechanism of rule (i.e. computation) and (associative) memory. In the most general understanding the main differences between rule-governed processes and analogy-governed processes can be formulated as follows. Rule-governed processes are derived by pure concatenation, and are characterized by being productive, applied to nonce words, and compositional in meaning. Analogy-governed processes have relatively low productivity and more semantic idiosyncrasy than rule-governed processes. In analogical formations the relationship between the overall meaning of an expression and the combined contribution of its constituent parts is not fully specified. Analogical word-formation is also very much limited in its extension to new coinage and is supported by associative memory that links the base forms and the derived units in the lexicon.

It is commonly assumed that the formation of some compounds should be rule-based whereas the formation of others should be analogy-based. This brings up the necessity of distinction between deterministic symbolic rules that underlie the formation of new coinages and pattern imitation, when a morphological structure emerges from the statistical regularities that characterise the forms and meanings of words (Hay and Baayen 2005). Following Booij (2007), this research assumes that the absolute distinction between the two is probably impossible since such rules should be viewed as strengthened cases of analogy and that a clear-cut line between the two is difficult to draw (Booij 2007, 2011, Krott 2009).

Some of the points put forward in a Construction Morphology framework (Booij 2005, 2007, 2010, 2011) are helpful in discussing the productivity patterns of nominal compounding in English. In this framework it is suggested that patterns of word formation are abstract schemas that speakers of the language use in the processes of producing morphologically complex sequences. These schemas arise in the process of generalising over sets of existing complex words that consistently demonstrate correlation between form and meaning and are used by speakers as templates for

[79] These include studies on: verbal inflection (Skousen 1989, Derwing & Skousen 1994, Albright and Hayes 2003, Keuleers 2008), the history of some negative prefixes (Chapman & Skousen 2005), compound stress (Plag et al. 2007, Arndt-Lappe 2011, Arndt-Lappe & Bell 2012), the choice between *-ity* and *-ness* (Arndt-Lappe 2012), adjectival comparison (Elzinga 2006).

coining new words, thus serving as bases for analogy. Booij (2005, 2007, 2011) demonstrates this by showing how a number of compound-initial words in Dutch are used not as constituents but in the way that affixes are normally used, by providing a number of examples in which the initial elements represent cases of the so-called 'semantic concentration', which means that the meaning of a word is limited to only one aspect of meaning. Words that have become concentrated only on one aspect of meaning (when used as initial elements in compound sequences) can develop a life of their own, and the connection with their original use can become obscure. For example, a word like *hoofd* (head) when used as a modifier of an N+N compound projects the meaning 'main'; the words like *dol* (mad), *beer* (bear), *bloed* (blood) function as intensifying prefixes; the word *scharrel* (to scratch, to scrape) is used in the combinations to convey the meaning 'free range' (examples from Booij 2007: 36). Booij (2007) considers the formation of new sequences that utilise the same pattern cases of analogy in which analogical patterns have transformed into constructional idioms.

As claimed by Jackendoff (2002), constructional idioms are morphological or syntactic schemas with one or more positions being lexically fixed. Other positions are viewed as open slots, represented by variables. If we think about modifiers in nominal compounds, once a word becomes part of an abstract schema it loses its morphological and semantic characteristics, which makes it somewhat similar to a prefix. However, this transformation is not full, since our language experience with the word suggests that it still belongs to a certain category and can be related to a certain lexeme.

The loss of noun characteristics by a modifier in an N+N sequence is clearly seen on the level of structural representation with the grammatical limits of the modifier component. Moreover, in the case with some modifier elements (e.g. *animal, future*) the morphological status of the modifier is unclear, when the words can be categorised as adjectives or nouns. However, there does not seem to be any ambiguity in terms of the words' grammar when they occur in other constructions. This means that on the level of N+N combinations the use of such words in the modifier position can be said to be idiomatic on the constructional level, and combinations like *animal + N* and *future + N* may to a point be considered constructional idioms.

Constructional idioms are understood as patterns with productivity, which (for compounds) means that one of the positions for a word is lexically specified. This makes it possible to avoid the necessity of deciding on the status (affixes or words) of initial words in compounds because their specifications, in terms of use and meaning, are registered in the constructional idioms (Booij

2007: 36). The results of the analysis carried out in the current research in general support this idea on the level of structural representation, with 95% of cases demonstrating the preference for being used in one position. However, the degree of preference differs, which implies that analogy in the formation is not always strengthened into a constructional idiom.

2.8.2 Analogy on the semantic level

In the cognitive framework analogy can be defined as:

> [a] general cognitive process that transfers specific information or knowledge from one instance or domain (the analogue, base, or source) to another (the target). Sets of percepts, whether visual images, auditory signals, experiences, or dreams, are compared, and higher-order generalisations are extracted and carried over to new sets (Blevins and Blevins 2009: 2).

It is obvious from this definition that analogy is not limited to the structural patterns but also involves other levels, e.g. level of conceptual representation.

Analogy in derivation can be seen in affixes that are used for deriving morphologically complex words. Affixes usually alter the meaning of a derivational base in a specifically prescribed way. Extending this to compounds, Booij (2007, 2010) claims that analogical compounding is based on an individual compound (model word) with an idiosyncratic meaning. This must be known for new compounds formed by analogy from the model compound to be understood. Analogy can be traced using semantic evidence when a particular idiosyncratic interpretation recurs in newly coined words (Booij 2007: 37). However, following the opinion expressed in Schlücker and Plag (2011: 1542), Krott, Schreuder, Baayen and Dressler (2007: 27), it is assumed here that analogy does not have to be driven by individual model words, but also involves paradigms that function as the basis for analogy. In this case not a single compound but a set, i.e. a paradigm (in our terms, constituent word family and its characteristics), predetermines the formation of new items. Krott, Schreuder, Baayen and Dressler (2007: 27-28) use the term 'paradigmatic analogy' and say that:

> In this type of analogy, the selection is based on the similarity of the target compound to a set (i.e., paradigm) of compounds, opposed to its similarity to a single exemplar, i.e., a single compound.

The studies to date, including the study on the use of linking elements in Dutch and German (Krott et al. 2007), the studies on stress assignment in compounds (Plag, Kunter and Lappe 2007, Plag 2010), and the studies on the interpretation of the semantic relations (Gagné and Shoben 1997, Gagné 2001), provide sufficient evidence that a paradigmatic approach to analogy provides a good account for the patterns noted in the formation of compounds. The analogical approach claims that the formation of new complex lexemes is based on the paradigms of similar existing complex

lexemes and their formal properties rather than on abstract rules (Schlücker and Plag 2011: 1540). Therefore, the form of new compound coinages relies on the formal and semantic properties that the constituent words in the new combinations share with other compounds these constituents occur in.

This thesis attempts to draw a parallel between the productivity patterns on the level of structural representation and on the semantic level. That is, the productivity of a constituent in a compound should be connected with the productivity of the semantic relation this constituent realises. This assumption is based on the idea of analogy, which should be revealed not only on the level of structural representation but should also find its reflection on the semantic level.

In the course of the analysis of the literature, I have found that most works on the use of analogy in producing and interpreting N+N compounds seem to look into the interpretational patterns of compounds. It has been often suggested that the modifier element has more weight in determining the relational reading a compound receives (Gagné 2001, Gagné and Shoben 1997, 2002, Gagné and Spalding 2004, van Jaarsveld, Coolen and Schreuder 1994, Storms and Wisniewski 2005). The psycholinguistic evidence obtained by Gagné and Shoben (1997, 2002) in the course of their studies suggests that the selection of a relation for a novel compound is affected by the speakers' linguistic and statistical knowledge about how the modifier is generally used in the language, which is directly connected to the modifier constituent word family (CWF) size. The same was noticed for compounds in German and Dutch. Krott (2009) claims that the modifier family appears to be more influential than the head family and its influence is clearly seen in the choice of linking elements, which is claimed to be determined by the semantics of the modifier rather than the semantics of the head. Based on this, Krott (2009) assumes that the semantic properties of the modifier element are more valuable for compound processing and are more important for coining new compound items. Moreover, this importance is attributed to the modifier, not to the first element in the compound. Storms and Wisniewski (2005) performed a psycholinguistic study of compounds in Indonesian, a language in which the modifier constituent follows the head. Their results suggest that the position of the modifier does not influence compound processing. This means that the subcategorisation process and meaning construal takes place on the conceptual level and may not be connected with the structural peculiarities of the language.

The head CWFs are reported not to be significant for processing compounds. As Gagné and Shoben (2002) comment, in the course of their study, the information from the head CWFs was

utilised by the participants only when they had to deal with ambiguous combinations. This implies that the properties of the head and the language speaker's experience with the head CWF are important for pinpointing the best possible interpretation for a given item. In my opinion, this fact seems to be underestimated, since it is evident that the importance of the head in dealing with ambiguity is crucial for understanding the nature of compounds and the way speakers operate with them. In fact, ambiguity of compounds is one of their specific characteristic features and in the course of the research reported in this thesis it has been observed that the semantic relations in the majority of compounds taken from real language situations are difficult to define unambiguously. This suggests that the process of their interpretation requires the use of information stored in the concept denoted by the head element.

Although compounding is a productive morphological process in English, the number of compounds that have a very high frequency of occurrence (or are institutionalised or lexicalised) is limited. A number of corpus studies that utilise different corpora, e.g. BNC, WordNet, CELEX (Andrews, Miller and Rayner 2004, Plag 2006, Plag et al. 2007, etc.), demonstrate that compound words generally tend to be very low-frequency. Since speakers are not often presented with one and the same item, it may be the case that the meaning of low frequency compounds is re-created rather than automatically accessed each time the speaker encounters them. As is outlined in a number of works (e.g. Gagné and Spalding 2009, Juhasz et al. 2005, Schreuder and Baayen 1995, 1997) processing of compounds involves constituent integration. This means that a compound is decomposed into its constituents during processing. Composition occurs on every level beginning from parsing a compound into structurally discrete elements, then goes on to semantic composition, when the meaning of compounds is constructed by integrating the semantic representations of the constituents.

It is obvious that the head and modifier components influence the semantics of the whole coinage differently. A functional analysis of the different roles performed by the modifier and the head noun can be of value in explaining this. The head noun denotes the general class, whereas the modifier specifies the subclass. So the modifier narrows the basic category. For example, in the compound *coffee table* the modifier narrows the basic category TABLE to allow for the reference to the object named by the compound. A different modifier, say *poker*, narrows the basic category TABLE to a different subcategory based on a different criterion thus providing a reference to a different object. Thus, it is the modifier that distinguishes the intended referent from other potential referents (by means of hyponymy, see Section 2.5.2 for detail). Principles of referential

communication emphasise the modifier. However, the question whether the realised relational information is dependent on the semantics of the head or the modifier remains unanswered.

The arguments above suggest that the role of the head in the formation and processing of meaning of N+N sequences is somewhat misinterpreted or underestimated. The importance of the constituents in terms of their semantic weight in compounds, in terms of profiling the relational information, and in terms of triggering analogy for coining new items using the same pattern cannot be compared since the functions they fulfil are inherently different. It is suggested here that the abstract schema of the head and the head's valence structure are responsible for the choice of the lexical item it can be compounded with. This suggestion is based on the assumption that the relational information responsible for the meaning conveyed by compounds is part of the nature of compounds. The ability to establish the relations between objects in extralinguistic reality and to transfer them into language is one of the basic human cognitive abilities that are developed from the early stages of perceiving the extralinguistic reality and coding the information by means of language. Years of practicing this ability transform it into an automatic skill.

Based on the discussion above, a number of points important for the current research have emerged. Firstly, the choice of particular nouns for N+N compounds is determined (but not entirely limited) by the semantic properties of the constituents. Combining two nouns into a conceptual combination is viewed as the result of the match of the valence structures of the modifier and head constituents. As compared to verbs, the valence structure of nouns is more flexible and allows for more freedom; therefore, theoretically, any combination is possible as long as it satisfies the pragmatic needs that arise in the course of communication. For example, nothing would prevent a speaker from coining a compound *bicycle rose* if there were an entity which could be thus labelled.

It is assumed here that the valence structure is the same for every noun but certain valences have a higher possibility for being realised depending on the noun's semantics and use in the language. For example, boxes are often used as containers; therefore, the corresponding valence has a higher probability of being activated in compounds with the constituent *box*. Therefore, I argue that based on the semantics of the noun concept it is possible to predict which valences are more likely to be activated when a noun concept is used in a conceptual combination. This suggestion is based on the results obtained in the course of studies by Maguire and colleagues (Maguire, Maguire and Cater 2010, Maguire, Wisniewski and Storms 2010), who demonstrated the

frequency with which a given relation is used with a given constituent is dependent on the semantic content of the given concept. For example, the concept LEATHER usually acts as the material/substance of which other objects consist, e.g. *leather jacket, leather bag, leather wallet*, and thus is more likely to occur in combination with concepts that can be made of leather. So, if we integrate the concept LEATHER with some other concept, then the COMPOSITION relation is most likely to be activated. However, this does not exclude the possibility of other valences to be activated, as can be seen in examples like *leather needle* where the relation between the constituents is not that of COMPOSITION. Although in our world needles are not normally made of leather, speakers still have no difficulty understanding the combination *leather needle*, because the semantic content of the constituent *needle* constrains the activation of the most frequent relation which may not make sense on the one hand, and allows for inferring of a relation that is sensible in this case, on the other hand. Thus, the semantic content of a combined concept together with our experience of this concept are believed to be responsible for constraining activation of all possible valences. As a result, different valences that a concept is claimed to possess have different probabilities for being activated.

The semantic relations that are inferred between the constituents of compounds are different from valences, with valences being potentially present in the semantic structure of the concept and relations being the result of the activation of this potential. I also argue that activation of semantic information by a noun in an N+N sequence is determined by the position it takes and the role it fulfils. The current research assumes the leading position of the head in the process of forming a conceptual combination.

The theoretical prerequisites outlined above will prove to be helpful in the investigation of two issues that are believed to be important for the study of nominal compounds, especially for the discussion of how compounds are produced and what mechanisms are involved.

The first issue concerns structural matters in compound formation. A noun constituent in a compound is expected to demonstrate a preference for being used either as a head or as a modifier. This expectation is based on Baayen's (2010) claim that the constituents of lexicalised compounds in English have a tendency to be used predominantly either as a head or as a modifier. The current study attempts to investigate whether this claim can be extended to all compounds irrespective of their status as lexicalised or not. Some studies claim that nouns are used equally productively in modifier and head roles, based on the occurrence of nouns in the head and modifier position in the BNC (e.g. Maguire, Wisniewski and Storms 2010). The data presented in

Maguire et al. (2010: 61) shows the proportion of noun occurrences in combination by modifier and head. The data is grouped in semantic categories that constituents belong to, e.g. Act, Animal, Location, Substance, etc. A superficial consideration of the data provides evidence to think that nouns that relate to certain semantic groups may demonstrate preference for being used in one position and not the other. For example, nouns that belong to the categories like Substance, Plant, Location and etc. seem to be used mainly as modifiers. Maguire et al. (2010) do not elaborate on this matter since it is not the purpose of their analysis.

The current research is aimed at looking into this matter in more detail in order to see whether compound constituents are preferably used in one role rather than the other and (if so) to discuss the factors that can be of importance.

The next issue that the current research is aimed to investigate concerns the connection between productivity of the constituent on the structural level and productivity of the semantic relation realised in this constituent family. The study by Maguire et al. (2010) mentioned above provides evidence that the semantic content of a concept strongly influences the way it is used in conceptual combinations. For example, combinations that fall into SUBSTANCE – ARTEFACT type are more likely to realise the MADE OF relation than the IS relation because artefacts are expected to have a constitution which can be specialised by concepts naming substances like *plastic, chocolate, wood*. As mentioned above, Maguire et al. (2010) mainly concentrate on the semantics of the constituents and come to the conclusion that similar modifiers and heads combine in similar ways. This implies that the pairings of modifiers and heads are not distributed randomly but fall into a number of regular semantic patterns which may reflect productive use of semantic relations. This finding has important implications for the current research.

Based on their analysis, Maguire et al. (2010) put forward and justify the claim that in the course of processing the meaning of a compound, we utilise our inner statistical knowledge about how nouns tend to be used in combination in order to facilitate the interpretation of novel compounds. This knowledge concerns our experience with a constituent, i.e. how a constituent (or semantically similar constituents) has been used in the past. This information is believed by the authors to be important for meaning construal in both formation and understanding of compound structures.

The above claim also allows for a suggestion that semantics of the concept can be an influential factor for predicting the way the constituent is likely to be used not only in terms of semantic collocation but also in terms of its position in a compound. My analysis of meaning of N+N compounds is based on the distinction that I make in this thesis between 'semantic content'

(dictionary information) and 'peripheral knowledge' (encyclopaedic, pragmatic information). This distinction is opposite to a Cognitive Linguistics view of meaning, where meaning in the mind is more than what is in the dictionary, and where no distinction is made between semantics and pragmatics. I view such a distinction, however, as a useful practice for my analysis, since I believe that semantics alone does not always allow for a sufficient explanation of the data and that we need to resort to broader knowledge about the concept (the semantic pole of the symbolic unit) to account for the data, which I attempt to show in Chapter 4. In the course of the analysis I use this distinction to also demonstrate to what degree the semantic content of a constituent of N+N compound can be used for the analysis of the semantics of the whole compounded structure.

Finally, based on the above, we can speculate that if there is a connection between the semantics of the noun concept and its productivity in one position, then we should be able to see this connection in the way this concept combines with other concepts to form N+N combinations. Since the semantic relation between the constituents is responsible for sense creation in a compound, the knowledge about how a given constituent contributes to the overall meaning should be part of the statistical knowledge mentioned above. Utilisation of this knowledge should result in the productivity of the semantic relation that this concept uses in a conceptual combination. So, if the constituent of a compound is used productively in one position, we should expect it also to favour one semantic relation over the others when used in this position.

2.9 Summary

In this chapter I provided a brief outline of the relevant (for this research) approaches to the semantics of N+N endocentric root compounds in English. I also looked at the issues of productivity in the formation of nominal compounds and discussed the possibility of the connection between the productivity of one constituent on the structural level and the productivity of the semantic relation this constituent realises in a conceptual combination. The realisation of the semantic relation by a constituent is suggested to be connected with the valence structure of the noun concept. Valences are understood as the potential of a noun to realise relational information in collocations, and semantic relations are viewed as physical realisations of this potential.

Based on the discussion of issues concerning nominal compounding in English, I believe that some of the principles of analysis and interpretation of the obtained results used in Ryder's (1994) approach may help in shedding light on the questions interesting for the current investigation. As pointed out above, this research is concerned with the way compounds are coined and also looks into other issues associated with the formation of N+N sequences, including:

- Can we talk about productive use of a noun as one of the constituents of an N+N sequence but not the other? What factors can be of importance in this respect?
- Is it possible to say that one of the constituents (and not the other) of an N+N compound is responsible for the relational reading a compound receives? If so, which of the constituents it is?
- Is it possible that the productive use of a constituent in the head or modifier paradigm coincides with the productive use of a semantic relation typical of this constituent in the larger paradigm?

These questions as well as other suggestions put forward in this chapter are checked and discussed in Chapters 3-5. I use the data collected from the BNC for drawing conclusions about the regularities for nouns used in compounds. Following Ó Séaghdha (2008), I believe that the analysis of the information from large corpora is helpful for making judgements and approximations concerning the kinds of conceptual knowledge involved in inferring meaning into linguistic structures. Bearing in mind that making judgements based on the corpus data may seem a poor proxy for knowledge gained through experience of the world, I still tend to agree with Ó

Séaghdha's (2008: 12) opinion that the analysis of corpus data can provide valuable insights into the issue of compound formation.

3. Methodology and initial results. Introductory comments.

The aim of this chapter is to present the methodology of the data collection and analysis. In order to investigate the issues outlined in Section 2.6, i.e. the issue of productivity of a noun as a constituent of a compound and the issue of productivity of the semantic relation activated by this constituent, a database of constituent word families (CWFs)[80] was collected. This chapter provides information on the stages of the data collection process and limitations of the search, and presents the details of the statistical analysis that was performed to check the general trends in relation to the issues above. It also explains the actions that were taken in order to see how the noted general trends surface on the level of individual constituent families.

3.1 Procedure

The data collection process consisted of several stages.

Stage 1 involved the collecting of a small sample of compounds and their semantic analysis. The purpose of collecting this small sample corpus was to see whether the chosen classification (Levi 1978)[81] could be used for a wide variety of combinations from the corpus and could then be used for further analysis. In a second part of the research, this small sample served as a source for a selection of 50 compounds, the distribution of whose elements was examined in greater detail as described in the next paragraph.

In Stage 2 I chose 50 compounds from the sample of 500 items collected during Stage 1. The 100 nouns that comprise the chosen 50 compounds served as the basis for collecting CWFs in order to check the productivity patterns for constituents of compounds. The semantic relations between the constituents of all compounds were analysed using Levi's (1978) classification.

Stage 3 involved statistical analyses of the data obtained in Stage 2.

In Stage 4 I was looking at individual constituent families to see how the noted trends are reflected on the level of a single constituent.

[80] The paradigms of compounds that share either the modifier or the head constituent are normally referred to as 'modifier family' and 'head family' respectively. The set of both families for one and the same noun is referred to as 'constituent word family' (CWF) (Krott et al. 2007, Schlücker and Plag 2011).

[81] Levi's (1978) set of semantic primitives is widely used as the basis of the research on the semantics of nominal compounds, due to its compactness, a suitable degree of generality and ability to cover big sets of data. The choice of this classification over the others is dictated by these reasons.

3.2 Stages of data collection and analysis

3.2.1 Stage 1

500 examples of nominal N+N compounds were picked from New Zealand printed media.[82] The examples had to satisfy the following criteria.

a) The examples were picked from the main body of articles; headlines were excluded.
b) The topics included business and economics, crime, social issues, art and entertainment, popular science.
c) The chosen examples had to demonstrate a relatively high degree of transparency, and be semantically analysable in the sense that it is possible to define a relation that connects the components. Exocentric compounds (e.g. *hatchback* and *lion heart*) and co-compounds (e.g. *fighter-bomber*, *singer-songwriter*) were not included in the data set because the original research questions involve only endocentric compounds. Since one of the criteria for collecting the data was semantic transparency, the majority of creative (metaphor-based) compounds like *jailbird* or *fleabag* were excluded from the analysis.
d) The collected compounds had to be endocentric.
e) Such criteria as stress, spelling, institutionalisation and lexicalisation were not taken into account. Lexicalised items like *party animal* and *lugsail* were excluded from the sample if they did not demonstrate a sufficient degree of semantic transparency.

The 500 examples were analysed and classified in accordance with Levi's (1978) set of semantic primitives.[83] The purpose of the analysis is not to define the relations between the constituents as precisely as possible (since the interpretation of the meaning of a compound is relatively subjective), but to show that similar patterns recur and suggest the reasons why this might be the case.

Semantically opaque situational formations like *lemon lady* were also avoided, since their meaning sometimes can only be pinpointed based on the situational context they occur in, and the discourse context might not be of much help (as in the example below).

> Cultured gents, identified for me by Mackie as being the Himalayan peaks of the Jockey Club, paid compliments to Tremayne from an adjacent table and bowed low to the sponsor. He, the **lemon lady**'s husband, eulogised Tremayne, who winced only slightly over Top Spin Lob being slurred to Topsy Blob,

[82] For the full list see Appendix 1.
[83] See Table 2.1 in Section 2.3.3.

and a minion in the livery of Castle Houses brought forth a tray bearing the award itself, a silver bowl rimmed by a circle of small galloping horses, an award actually worthy of the occasion.

It is necessary to point out that the constraints were imposed to make the task of analysing very complex data relating to compounds manageable. There is no particular reason to expect different results if the above constraints were not observed, although the possibility of different results of the analysis should not be ruled out. The constituents of compounds from the corpus were then checked for their frequency of occurrence in the BNC, and the ones whose frequency was 1000 or more occurrences per 100,000,000 were considered candidates that could form the basis for collecting a corpus of constituent word families. The frequency limitation was set in order to be able to collect a corpus of compounds in which a given word may occur either in the head or modifier position.

Overall, the results of the semantic analysis of the relations connecting the elements of compounds in the collected corpus showed that the distribution of the relations in compounds is not even, and some of the categories occur much more often than others. Almost two thirds of all the compounds in the collected data set realised only three semantic relations: purpose (34%), possession (16.2%) and spatial (14.4%). Such uneven distribution of the semantic categories is assumed to be natural for English compounds in general, and certain categories seem to occur more frequently in certain nominal constructions (Bauer and Tarasova to appear). It was initially planned to choose 50 random compounds for further analysis, but such an uneven distribution of semantic relations required stratification of the sample in order to include compounds from different categories.[84]

3.2.2 Stage 2

50 compounds were chosen from the sample of 500 items for the corpus experiment. In order to obtain a set of compounds that is representative of the natural distribution of semantic relations in English N+N compounds, the number of items in each semantic category was proportionate to the number of items that occur in the initially collected sample of 500 compounds (Stage 1). Another requirement was that each of the 100 nouns in the 50 compounds had to be different.

100 different constituents comprising these 50 items served as the basis for extracting the paradigms of compounds (constituent word families – CWFs) in which each of the constituents

[84] Proportionate allocation strategy was used; i.e. a sampling fraction (the ratio of the sample size to the size of the stratum) was picked in each of the strata that was proportional to that of the total corpus.

occurs. The semantic relations realised in compounds comprising CWFs were determined based on Levi's (1978) classification.

The semantic relations in the compounds have a number marker, which signifies the directionality of the relation. Levi (1978: 76) shows that the syntactic structure of the interpretation of at least three of the semantic relations (CAUSE, HAVE, MAKE) demonstrates bidirectionality, e.g. *tear gas* – 'gas CAUSE tears' vs. *drug deaths* – 'drug(s) CAUSE deaths'. In the collected corpus the bidirectionality of the relations is more widespread across the categories and includes LOCATION (*office building* vs. *garden buildings*), SOURCE (*photon energy* vs. *life energy*), ESSIVE (*mansion house* vs. *tower house*), TIME[85] (*crisis year* vs. *crisis decisions*), INSTRUMENT (*farm machinery* vs. *wind farm*). In the process of picking the sample the distinction into the different directions of a relation was not made, since for some categories the number of nouns was not large enough to be included in the stratified sample. However, some recent works in compounding (Plag et al. 2007, Kunter 2011), as well as our analysis of the semantics of constituent word families, demonstrate the necessity to extend Levi's list in order to make a distinction in the direction of the relation.

The list of compounds that were chosen for collecting the corpus of constituent word families is given in Appendix 2.

As previously mentioned, semantic relations between the components of a compound may be ambiguous, and one and the same compound may be referred to different categories by different researchers. In order to reduce subjectivity, three raters were asked to provide their interpretations for the chosen set of 50 compounds. All the raters were native speakers of English, PhD students in Linguistics who work in different areas, and their spheres of expertise do not include morphology, semantics or nominal compounding. The raters were provided with some background information and a table taken from Levi (1978: 76) and were instructed to think about the given compounds in their own time and classify them according to Levi's (1978) categories. The results obtained from the raters are presented in the table (Appendix 3). Overall the results are considered to be satisfactory and no compounds were excluded from further analysis.

The set of 50 compounds was then used as the basis for collecting a constituent word family corpus. Each of 100 nouns that constitute the compounds was put into the search engine of *BYU-BNC: The British National Corpus* (Davies 2004-) in order to collect N+N combinations a given noun occurs in. Each noun was put into the SEARCH field and [nn*] was typed into the COLLOCATES

[85] The necessity to account for the distinction between the TIME and LOCATION relational concepts is discussed in Section 2.3.3.

field. The collocations were limited by one noun before and one noun after the searched item. Then, where there was one, the plural form of the same noun was put into the search engine in order not to miss compounds in which the searched constituent in the head position does not occur in the BNC in the singular form.

3.2.3 Limitations on the search

Some recent research (Štekauer 2005, Gagne and Spalding 2010, Maguire et al. 2010) suggests that the meaning realised in compounds can be predicted. The massive corpus study by Maguire et al. (2010) has shown that the semantic content of a concept strongly influences how it is used in combination with other words, and generalised information regarding the semantic categories of the modifier and the head can be useful in diagnosing the relation between them. This information is related to the statistical knowledge individuals possess about the way words can be used in combinations. The analysis carried out in the current research was intended to show the tendency noted in Maguire et al.'s (2010) study; however, the task was approached differently. The search of the BNC database was targeted and the examples were hand-picked in order to avoid the combinations that would not satisfy the criteria, as in examples like *century house* and *umbrella back*, which either have a different bracketing structure (as in the first case) or are not compounds at all.

> … so we found a minister who was very excited by the idea of marrying us in a seventeen [sic] **century house** , and in fact, on the morning of the wedding he was more worried about what he was wearing than what I was!
>
> Give me my **umbrella back**.

Since I was interested in the number of types of compounds rather than the number of tokens, there were no limitations concerning the frequency of the collected items. Moreover, compound hapaxes are generally believed to provide important information about the development of the morphological pattern at the current time (Baayen 1992, 1993, Baayen and Lieber 1991, Ricca 2010).

In my study there was another reason for manual picking of the examples, which is the necessity to exclude so-called synthetic compounds. It does not seem possible to set the limitations for these using the corpus tools, as the morphological criterion offered by Lieber (1992) does not always work. For instance, if we set the limitations on nouns with the agentive endings *-er* and *-or* in the head position, then compounds with head words like *professor* and *minister* will be excluded from the analysis. In the current research I do not follow Lieber's (1992) suggestion that

all compounds in which the head element is a deverbal noun should be considered synthetic. In this research compounds are considered synthetic only if the modifier functions as a direct object of the derivational base of the head noun in the interpretation, e.g. *literature teacher* – teach literature, *company management* – manage a company. Combinations like *art competition, tourist information, state prosecutor, weight division*, etc. are not considered to be synthetic since they require relational interpretations. The so-called NDVCs[86] (non-affixal deverbal compounds) like *history research*, *business talk* are included in the corpus if their interpretation involves one of Levi's (1978) semantic relations.

Each computer search of the corpus produced a list of items that a noun collocates with, each of which was checked against the limitations criteria. Despite the limitations for N+N combinations, the surrounding context of the search item was also consulted in order to exclude cases that do not satisfy the search criteria. Some of the examples of the latter are shown below.

> This is the **family run** Oakes clothes shop in Ledbury.
>
> In other words, erm that er special **landscape areas countryside** is not to change quite radically.
>
> That is exactly what he did -- and at 7.30pm on **Wednesday night police** first learned of the daring raid.

As a result of the data collection 7332 compounds that make up 197 head and modifier constituent word families[87] were collected for further analysis. The semantic relations realised in the collected N+N sequences were analysed and the number of occurrences of each of the compounds in a given constituent word family was checked (see Appendix 4).

3.2.4 Stage 3

The third stage of the data collection involves statistical analysis in order to obtain a picture of the general trends.

The statistical analysis of the data was performed to test the following hypotheses:

(1) There is a preference for a noun to be used more productively in only one position (modifier or head) but not the other.
(2) The constituent's preference for one position on the level of structural representation correlates with its being used in the same way semantically within the constituent's

[86] The term NDVC is offered in Lieber (2010).
[87] Although the expected number of constituent word families is 200, three of the nouns (*matter, future, figure*) appear only in one position in N+N combinations in the collected corpus. These were still included in the statistical analysis; however, in order to carry out the tests, one was added to all family sizes.

modifier or head paradigm; and there is a connection between the family size and the constituent's concentration on one semantic relation.

I will look at the hypotheses listed above one by one in the course of my discussion.

The first hypothesis tested through statistical analysis was aimed at answering the question of whether a compound constituent is used more productively in one position (modifier or head) but not the other.

In order to test this the collected corpus data were coded on the basis of a table that listed for each compound constituent the number corresponding to the family size of this constituent as it is used in the head position (famSizeN2) and in the modifier position (famSizeN1).[88] Table 3.1 gives an overview of the distribution of family sizes. Family sizes of 0 occurred only once in N2 position, and only twice in N1 position (with each position having 100 different nouns).

Table 3.1 Distribution of family sizes

Constituent	minimum	maximum	mean	standard deviation
N1	0	83	31.79	20.47
N2	0	116	45.95	30.26

To reduce the skewing of the distribution of the family sizes, and in order to alleviate possibly harmful effects of extreme values on our statistical models, the family sizes were log transformed (see Baayen 2008: 38). In doing so we also follow the standard practice with compound family size measures in other studies (e.g. Bell and Plag 2012). Since some of the frequencies were 0, we added 1 to all family sizes before log transformation. The statistical analysis at this stage was carried out using the statistical package R (Bates, Maechler and Dai 2007, Pinheiro and Bates 2000, R Development Core Team 2011).

In order to test hypothesis (1) we devised a linear regression model with the log family size of N1 as the independent variable and the log family size of N2 as the dependent variable. The model with the best fit included a quadratic term for the independent variable and showed highly significant correlations between the terms for family size of N1 and family size of N2. Table 3.2 gives the coefficients of the model.

[88] The coded data table is presented in Appendix 5.

Table 3.2 Model coefficients for linear model showing the relationship between N1 and N2 family sizes. N1 = 98, Adjusted R^2 = 0.3796

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.43574	0.42932	8.003	2.65e-12 ***
logFamSizeN1	1.07457	0.32638	3.292	0.00139 **
$(logFamSizeN1)^2$	-0.30143	0.06094	-4.946	3.17e-06 ***

Figure 3.1 below is the graphic representation of the findings.

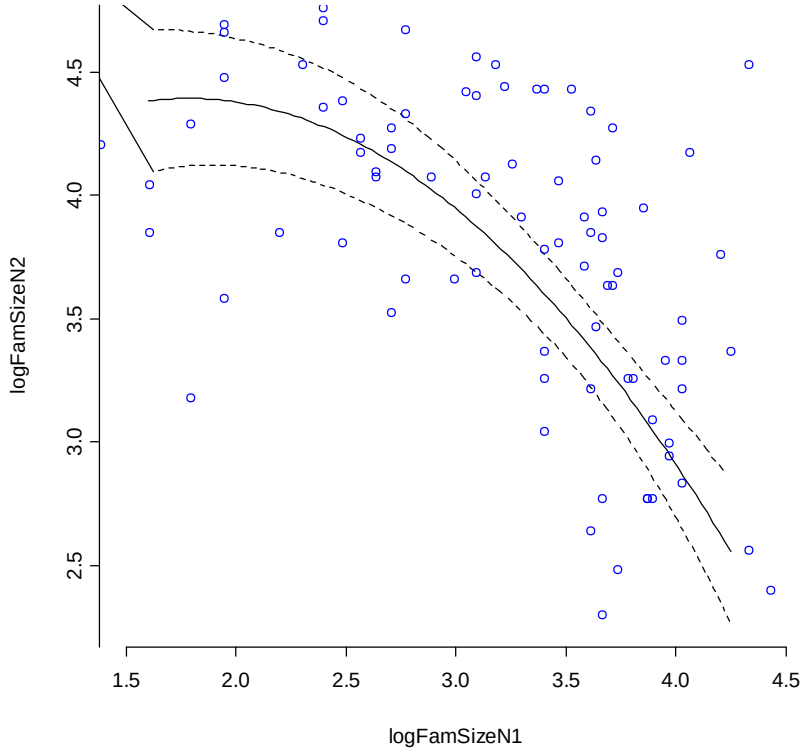

Figure 3.1 Relationship between family sizes.

The graph shows for each noun the size of its family in N1 position and in N2. The vertical and the horizontal axes represent the logs of family sizes for constituents used as heads (logFamsizeN2) and modifiers (logFamsizeN1) respectively. The dots represent the observed values for nouns that are attested both in N1 and N2 position. The solid line is the regression line, with the 95 percent confidence interval given by the two dotted lines. The negative coefficient for the quadratic term gives us the downward slope of the regression line (see, for example, Baayen 2008: Ch 4). As can be seen from the graph, the dots are not very concentrated. However, the distribution and the shape of the regression line clearly demonstrate that the larger the head family is for a given

noun, the smaller the modifier family is for the same noun. This means that the more often thae given noun occurs as a head of an N+N sequence, the less often it occurs as a modifier. As can be seen from Figure 3.1 this effect is not very pronounced for smaller N1 family size and becomes more pronounced with rising N1 family sizes.

Now we will turn to hypothesis (2), according to which we expect to see a connection between the family size and the constituent's concentration on one semantic relation.

Following the basic tenets of the slot-filling models (Murphy 1988, Wisniewski 1997), the current research is head-oriented and assumes the position that the semantics of the head noun serves as the schema whose properties are responsible for attracting certain modifiers based on the slots that are more available for filling. The availability of slots is determined firstly by the semantic properties of the head and only then by other factors, including our previous linguistic and nonlinguistic experience. In the case of modifiers the situation is the opposite and our previous linguistic and nonlinguistic experience of the modifying word is of primary importance.

Hypothesis (2), as given above, states:

> (2) The constituent's preference for one position on the level of structural representation correlates with its being used in the same way semantically within the constituent's modifier or head paradigm; and there is a connection between the family size and the constituent's concentration on one semantic relation.

In order to test it, I will look at the two suggestions that this hypothesis implies:

> (2a) The constituent should demonstrate a preference for certain relations, or even one particular relation, over other relations.
>
> (2b) This tendency should be stronger the larger the constituent family is.

So, if we suggest that there is a preference for one semantic relation over the others within a given constituent word family, then how can this suggestion be statistically measured? In order to do this, I devised a measure that I call 'instantiation index'. In order to determine the instantiation index the number of instantiations of different relations (DiffInstN1 and DiffInstN2) and the most frequent relation (HighestInstN1 and HighestInstN2) in each constituent family was calculated and added to the coded data table. Then the proportion of the highest instantiation among all family members was computed.

In order to additionally factor in the diversity of meanings present in the family, this proportion was multiplied with the number of different relations attested in the family. For example, if we

have a family size of 20 with eight different relations, and 10 compounds have the most frequent relation A, the index is (10/20)*8 = 4. If this family had only two different relations, the fact that 10 out of 20 have relation A is not so striking. This is exactly what we can see from the Instantiations Index, which will be much lower in this case: (10/20)* 2 = 1. So, we can assume that the higher the instantiations index is, the more concentrated the constituent family is on one relation.

In reference to the hypothesis in (2b), we should also expect the effect of the family size, i.e. the index should be higher for larger constituent families.

In order to test the hypotheses in (2a) and (2b) the instantiation indices for constituents N1 and N2 were calculated and used for the analysis. Since it made sense to check both (2a) and (2b) simultaneously, the obtained measures were used for testing the family size effect, as outlined in (2b). We fitted two linear regression models (one for each constituent) which would allow us to see if there is a connection between the family size and the constituent's concentration on one semantic relation. Log family size was used as an independent variable and log instantiation index as a dependent variable. Tables 3.3 and 3.4 below give the coefficients of the two models respectively.

Table 3.3 Model coefficients for linear model showing the relationship between family size N1 and instantiations of relations index for N1. N1 = 98, Adjusted R^2 = 0.6503

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	0.34443	0.08427	4.087	9.06e-05
logFamSizeN1	0.33816	0.02511	13.468	< 2e-16

Table 3.4 Model coefficients for linear model showing the relationship between family size N2 and instantiations of relations index for N2. N2 = 99, Adjusted R^2 = 0.4333

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	0.40984	0.11722	3.496	0.000713
logFamSizeN2	0.27787	0.03189	8.715	7.95e-14

As can be seen from the tables, the family size is a highly significant predictor in both models, which, together with the positive coefficients, supports hypothesis (2b) that the constituent's concentration on one semantic relation is stronger in larger constituent families. Note also that these models have a very good fit and can explain a large proportion of the variance (R^2 = 0.65 for N1, R^2 = 0.43 for N2).

Figures 2 and 3 below represent the findings in graphical form. The vertical axes in both graphs show the measures for instantiation indices of semantic relations in a constituent family, and the horizontal axes show the log measures for the respective family size. The lines are the regression lines of the respective models.

The lines show that with increasing family sizes, the instantiation indices become higher. And this means that compounds with larger family sizes tend to be more concentrated on one relation than compounds with smaller constituent family sizes.

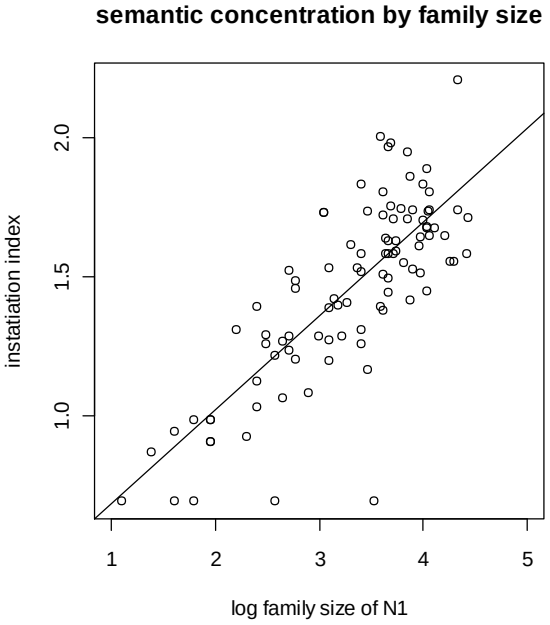

Figure 3.2 Relationship between N1 family size and the ratio of instantiations of different semantic relations

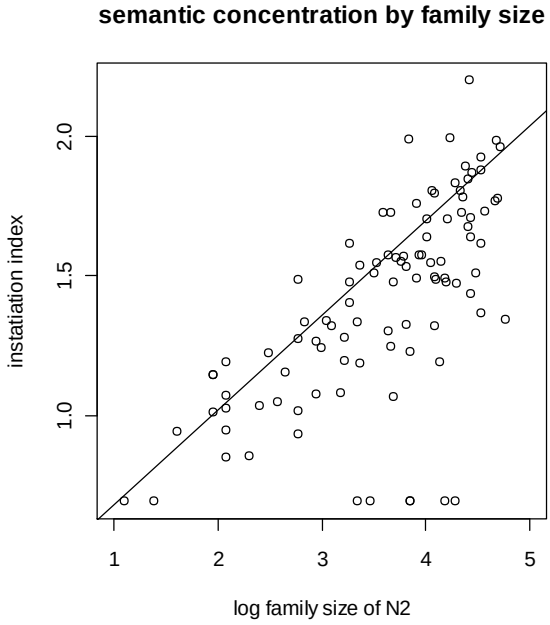

Figure 3.3 Relationship between N2 family size and the ratio of instantiations of different semantic relations

Thus, we have shown that there is a connection between the productive use of a compound constituent in one position and productive use of the semantic relation, which can be considered evidence of interaction between the levels of structural and semantic representation in compounds.

3.2.5 Stage 4

Since there is no methodology for defining the probability rate (PR)[89] for a given noun concept, weighting the PRs for the nouns in the collected corpus was performed *post factum* and was based on the results of the study. The valences whose frequency in a head or modifier compound family is 30% or more were assigned PR 1. This means that in an N+N sequence a noun is most likely to be connected with another noun via a certain relation. For example, in *family + N* compounds the

[89] See Section 2.5.3 for the discussion on probability rates of valences.

POSSESSION relation is realised in almost 40% of all cases, whereas other relations do not have such a high probability. Therefore, it is suggested that the valence with the highest probability rate (PR1) for the noun *family* used in the modifier position is that of POSSESSION. Probability rates 2 and 3 were assigned to valences that occurred in 11%-29% and 1%-10% of all cases in the constituent family respectively.

In order to account for factors that can be of importance for the noted trends, I looked at the individual constituent families. I used a Chi-Square test (*PASW Statistics software*, version 18) for each constituent. The test looks at the preference for realising one relation in each individual constituent family. 76 out of 100 constituents pass the test ($p < .001$). In 16 cases that did not produce satisfactory results, we seem to deal with exceptions rather than cases that fail the test. They are exclusions because the examples do not meet the requirements of the test. In eight cases (*mark, lemon, board, ship, police, credit, chocolate, secretary*) the number of compounds in one of the positions is too small (≤10), which hampered the results, since the number of compounds is important for the test. In eight other cases (e.g. with the constituent word families containing words *army, station, work, letter, bus, service, weight, technology*) one and the same relation appears to be the preferred one for modifier and head constituent families. However, the test can only show a positive result if the values are mutually exclusive. In the context of the current research this means that the same semantic relation should not have the same value in both positions, which does not always happen.

Taking the first two points into the consideration, the genuine 'fail cases' are the ones in which the distribution of categories is too broad and no relational preference can be seen. This is the case for only three constituent word families (that contain components *money, computer,* and *land*). As a result, no statistically valid preferences for realising one or two semantic relations can be seen in these constituent word families.

In five cases of constituent word families (that contain components *phone, plane, holiday, client, ship*) none of the above mentioned factors seems to be very salient. However, there could be a combination of two factors, i.e. the number of the compounds with a given element in a certain position is not very big and the preference for a relation in one position is not very strong.

As was shown by the analysis of the collected corpus, these deviations from the overall tendency for a constituent to prefer one relation over the others can be considered insignificant.

The next chapter looks into the separate constituent families and discusses the noted trends in more detail.

4. Results and discussion

4.1 Activation of valences by constituents: Introductory comments

One of the suggestions put forward in this research is that the set of valences nouns possess is the same for every noun and corresponds to the thematic relations[90] that are used in the linguistic analysis to label the way in which objects and phenomena of real life are connected. However, the potential for activation of individual valences differs from one noun concept to another. Since in this research the semantic relations connecting the constituents of compounds are viewed as the result of the activation of one of the valences from the set in the process of meaning construal, the analysis of the semantic relations in constituent word families is claimed to provide information on what valences have a higher PR (Probability Rate (Štekauer 2005), i.e. probability of being realised in an N+N sequence) for a given noun. In what follows, I will attempt to show this by looking at constituent word families in the collected corpus and will suggest which factors underline a tendency for a given noun to realise some valences and not others. In order to provide illustrative examples, I will be looking at the largest constituent word families in the collected corpus, namely containing the constituents *animal*, *community* and *business*.

It is also necessary to point out that some N+N sequences may be interpreted differently from what I offer below. The judgements concerning the interpretations of problematic cases were made based on the analysis of the nearest contexts the sequences are registered to occur.

The purpose of this chapter is to give an analysis of the compound data and see if the results obtained for individual constituent families are helpful in answering the questions outlined above.

4.1.1 animal + N, N + animal compounds: Miscellaneous notes

The constituent word family with the element *animal* comprises 66 items with the word being used in the modifier position and 42 items when it is used as a head.

According to the definition of the word in the Oxford dictionary (Oxford online 2012),[91] an animal is "a living organism which feeds on organic matter, typically having specialised sense organs and nervous system and able to respond rapidly to stimuli". Based on this, the following propositions are inferred: an animal lives, feeds, has senses, has a nervous system, responds to stimuli. These

[90] Following Estes, Golonka and Jones (to appear) thematic relations are understood as any temporal, spatial, causal, or functional relation between things.
[91] The list of links for online dictionaries is given in the Reference list.

propositions should give an idea of what valences may be potentially active for the concept denoted by the noun *animal*: lives where? (forest, farm, field, ocean); lives how? (pack, group); feeds on what (meat, grass, both); has what kind of senses? (smell, taste, touch). Animals are also opposed to other representatives of flora and fauna (e.g. fish, insects, reptiles, plants, etc.). Some features typical of animal habits and behaviour can be transferred onto humans by means of metaphor as in *party animal, political animal*.

4.1.2 animal + N compounds: The morphology of the modifier

The word *animal* is listed in dictionaries as a noun and as an adjective. In the latter case its registered meaning is usually that of 'relating to or characteristic of animals' (Oxford online 2012, Longman online 2012, Webster online 2012). However, there seems to be confusion about the morphological status of the word in dictionaries. For example, Longman (Longman online 2012) refers the word *animal* to nouns in combinations like *animal welfare* and *animal kingdom*, whereas Oxford (Oxford online 2012) provides the same examples to illustrate its use as an adjective.

So, the word *animal*'s being often used in the modifier position (according to the data, more often than in the head position) and fulfilling the attributive function is assumed to be responsible for the loss of characteristics typical of a noun and for it being perceived as an adjective by speakers of English. As an attribute in a given N+N combination, *animal* projects only some aspects of its meaning onto the head noun. This contributes to the loss of the linguistic characteristics of a noun; that is a partial loss of different aspects of meaning and a partial loss of grammatical characteristics. As a result the item *animal* has an unclear morphological status. How far can we claim that if a noun loses its noun-like features in an N+N sequence, it should necessarily become an adjective? Connor (2010: 34) points out that in many respects the decision about lexical entries taken by lexicographers is often guided by practicality and custom rather than by general rules of the language. Therefore, in this research, borderline cases, in which the morphological status of the modifier is ambiguous, are taken to be nouns if they are registered in dictionaries as nouns in their first meaning.

4.1.3 animal + N, N + animal compounds: Predictability of the relational information

Based on the definition, it is difficult to judge what the potentially active valences for the noun concept *animal* should be. The only ones that seem to be clearly deducible should be realised as the LOCATION relation as in *farm animal*, *island animal*, *zoo animal* and the POSSESSION relation as in *animal fur, animal instinct, animal body*. So, in the case with the noun *animal*, the dictionary entries which were considered do not seem to provide enough information about the semantic content of the noun concept but are more focussed on the encyclopaedic information. In order to fully understand the concept, we need to apply our common knowledge, which suggests that animals can be used as a source of food as in *meat animal, animal fat*; animals can be imitated in man-made works as in *ivory animal, glass animal*; they can form communities of a certain kind as in *animal population, animal kingdom*, etc. This kind of information comes from pragmatic knowledge and life experience, thus contributing to overall knowledge about the concept denoted by the noun *animal.* Although this information is not reflected in the dictionary definition, it is potentially present in the concept, and is also revealed in the constituent word families with the analysed element *animal.*

The statistical analysis of the data confirms the prediction that the valence realised as the LOCATION relation is prevalent in the constituent word family with the element *animal* used as the head and has the highest number of occurrences in this position (45.2%). The prevalent use of the LOCATION relation in *N + animal* combinations implies that one of the main ways for distinguishing between the hyponyms of the head element is through activation of the locative slot. The data suggests that conceptualising objects denoted by *N + animal* compounds via the realisation of the LOCATION relation is productive with the majority of combinations having a low frequency of occurrence in the BNC. The types of locations denoted by the modifier concepts range from natural habitats (land, forest, sea) to man-made ones (farmhouse, zoo, circus, laboratory). A closer look at the semantics of the locative *N + animal* compounds reveals that in the case with made-man objects as modifiers, the LOCATION relation has the shade of the PURPOSE meaning in it as well. This is not surprising, since the modifier nouns in this category denote objects that were created in order to fulfil a certain function, whereas the purpose of usage is not so explicit in nouns denoting natural objects.

The other valences that are active for the word *animal* in the head position correspond to the following relations: ESSIVE (*toy animal*), PURPOSE (*test animal*) (PR2); COMPOSITION (*glass animal*), SOURCE (*meat animal*), POSSESSION (*pedigree animal*) (PR3).

It can be seen in the above examples that the relational information connecting the elements is somewhat ambiguous and could have been interpreted differently. For example, the semantic relation in the compound *test animal* can be interpreted as INSTRUMENT, and the compound *meat animal* can be understood as 'animal bred FOR meat', thus providing the interpretation that projects the PURPOSE relation. I assume this to be the consequence of the overlap of the abstract relational concepts, which results in the ambiguity of the interpretation.

The relational information projected by the noun *animal* in the modifier position differs in different combinations. The array of relational meanings realised in *animal + N* compounds is quite large and the semantic analysis of the collected modifier word family does not reveal a preference for any valences that can be assigned the status of PR 1. The active valences realised as semantic relations are POSSESSION (*animal instinct*), PURPOSE (*animal care*), SOURCE (*animal protein*) – PR 2; COMPOSITION (*animal world*), ESSIVE (*animal resources*), INSTRUMENT (*animal tests*), TOPIC (*animal story*), LOCATION (*animal farm*) – PR 3. Such an array of relatively active valences suggests that the noun has a high relational potential in the modifier position. This may be explained by the word's functioning as a noun on the semantic level, despite the fact that it fulfils the function typical of an adjective. It might be the case that the relations in this group are totally determined by the semantics of the head. However, the meaning of the word *animal* in *animal + N* compounds is not similar in its semantic behaviour to Adj + N compounds because of the variety of the predicates implied in combinations. This argument can be used against the position taken in the dictionaries mentioned earlier, i.e. in *animal + N* sequences, *animal* is an adjective. However, the position taken by lexicographers is most likely based on the etymology of both words, since they entered the English lexicon separately and via different sources. According to the information in the Oxford Dictionaries (Oxford online 2012), the noun *animal* originates from Latin *animalis* 'having breath, from *anima* 'breath', whereas the adjective *animal* comes from the same Latin form but via French. So, it seems that different language sources and, probably, different times of borrowing give grounds to lexicologists for listing the adjectival and nominal forms as separate entries. But even if we agree with that, there appears to be a problem with classifying the adjective *animal* according to the existing classifications. Firstly, if *animal* is an adjective, it should be categorised as a nonpredicate one and is in many respects similar to the group that Giegerich

(2005) calls 'associative intersective adjectives', e.g. *bovine, feline, dental*. Adj + N combinations containing adjectives of this type do not fit into the gloss 'N is Adj', as, for example, compounds of the '*blackbird* type' do. However, in the case of associative adjectives there is always a noun that they are associated with (*cow, cat,* and *teeth*, respectively). This does not seem to be the case for the word *animal*, since the form of the adjective is homonymous to the form of the noun. Secondly, in *animal + N* combinations the modifier seems closest in terms of its syntactic and semantic behaviour to compounds of *social worker, mental health* and *civic centre* type, but still in the case of *animal + N* compounds, the existence of the homonymous noun makes it different from this type too. So, since there does not seem to be unambiguous evidence that would allow us to believe that *animal* in the modifier position should be treated as an adjective or as a noun, it is safe to assume that in *animal + N* compounds, the modifier may be closer to a noun than an adjective in its linguistic nature at the current moment of time, despite the fact that we are dealing with two etymologically distinct forms.

The realisation of a number of relations as part of the meaning construal of a compound noun is suggested to be a direct reflection of the way we perceive and position a given concept (and what stands behind it), in our case ANIMAL, in relation to other concepts in our mentality. Our knowledge about the concept, in its turn, is determined by what we know about the way different entities interact in the real world and what the features of such interactions are. This knowledge allows for the location of coinciding conceptual information and for the realisation of this information in language constructions.

If we look at the meaning realised by the noun in the modifier word family with the element *animal*, we can see that the concept ANIMAL is consistently used in its primary sense (a living organism) across the groups in which different relations are realised, except for the POSSESSION relation. The latter also comprises a group in which compounds may have a dual interpretation, literal and figurative. Among those are compounds like *animal power, animal passion*, etc. In these cases the figurative interpretation profiles people as possessing qualities typical of animals rather than humans, thus implying the opposition of the concept PEOPLE to the concept ANIMAL. The analysis of the contexts in which items of this kind occur has shown that although the figurative interpretation is preferable, they can be used in the literal meaning (cf. the examples from the BNC).

1 (a). One remains unassuming with a smack of humility, the other aggressive with a raw **animal power**.

1 (b).The rynd and the spindle soon came to be made from iron to be more durable, and the handle was developed to become a lever for turning the runner stone by **animal power**.

2 (a). Joely Richardson as Lady C is not as timid as Lawrence's confused heroine, while Mellors as played by Sean Bean is striking more for his designer courting britches than any **animal passion** .

2 (b). … united by our **animal passion**; Working so closely with our furry and feathered friends can be really rewarding... but you have to be able to deal with humans as well.

Thus, in this case the metaphorical transfer does not make these compounds opaque, but instead adds the necessity of involving associations in order to access the implied sense of 'humans possess N2 of the same quality as animals do'. Therefore the interpretation by means of the POSSESSION relation is preserved. However, the use of the word *animal* in this meaning gives grounds for a speculation that the metaphorical transfer contributes to the overall perception of the word in the modifier position as an adjective, which may also give grounds for the further development of the pattern of the word's being used as modifier more often than as a head.

4.1.4 animal + N, N + animal compounds: Asymmetry of the semantics

Returning to the question about the semantic value of the modifier or the head having in the construal of meaning of an N+N compound, it is important to look into the concepts expressed by the head nouns in *animal + N* compounds inside the relational groups. As suggested before, if the valence activated by one of the components in a compound (and not the other) occurs more frequently and in a larger number of sequences with one element than with the other one, then we can say that this component has more value in terms of contributing to the overall meaning of a compound. When discussing the head constituent word family with the noun *animal*, it was shown that, based on the collected data[92], firstly, in this position the realisation of one relation (LOCATION) is predominant and secondly, the semantics of the modifiers within different relational subgroups is more or less homogeneous, e.g. in the LOCATION group the modifier nouns denote a place or a habitat of some kind. This allows for an early suggestion that in *N + animal* compounds the activity of the locative valence is probably determined by the semantic structure of the head noun *animal*. This results in attracting locative nouns as modifiers in *N + animal* sequences. The locative meaning projected by the modifiers contributes to perceiving modifier nouns as dominating in meaning construal. Can this suggestion be extended to the *animal + N* compounds? It can be noticed that a number of head words within relational groups can be referred to the same semantic fields. For example, in the group that realises the COMPOSITION

[92] See Section 4.1.3 for detail.

relation the majority of head nouns have a common meaning that is associated with inhabitants of a particular area, e.g. *kingdom, world, community*. In the PURPOSE group of *animal + N* compounds the head nouns typically denote places, organisations and people that are aimed at providing for animals' needs (*animal hospital, animal charity, animal activists*). However, we cannot claim that the lists of head nouns within each relational group are homogeneous and that all head nouns bear strong features of relational similarity. For example, although the majority of nouns in the SOURCE group are from the same semantic field that contains words describing a living organism (*fats, cells, bones, insulin, tissues, flesh, skins, manure, noises, waste*), the sphere of usage of these words is slightly different. In the POSSESSION group, the head nouns can be divided into three distinct groups depending on whether the word *animal* is used in its literal or figurative sense. In the compounds characterised by the metaphoric transfer of animal qualities onto humans, the head nouns denote abstract qualities that are typical of humans rather than animals: *cruelty, passion, magnetism*. The second subgroup represents ambiguous sequences which can be considered metaphorical in some contexts and literal in others: *animal power, animal orientation*. The third group contains nouns that describe basic prerequisites of a living being: *body, head, physiology, behaviour, rights*. However, despite the differences in the semantics of the head nouns, the fact that the semantic relation realised in less homogeneous groups of head nouns suggests that in the case with *animal + N* compounds, the modifier plays a more important role in defining the reading a compound receives. This goes in line with the suggestion put forward in a number of works (Gagné and Shoben 1997, Gagné and Spalding 2004, Gagné and Spalding 2006a). It is also worth mentioning that, as discussed above, the relational information realised in *animal + N* sequences does not seem to be easily predictable from the semantics of the word *animal* as registered in dictionaries (if we take the definitions given in modern dictionaries as being a more or less plausible source of information that allows for considering a dictionary a close enough reflection of the semantic content of a word). It is possible to suggest that despite the seemingly obvious priority of the modifier in terms of the semantic value that is responsible for the relational reading of a compound, this information does not root in the semantic content of the modifier noun but requires elaboration of the meaning on the deeper level. Looking at the constituent word families with the element *animal* in the head or modifier positions, it appears that dictionary definitions are more helpful in predicting the meaning of *N + animal* compounds, rather than of *animal + N* compounds. This issue seems interesting and we will see if the same is true for other CWFs or whether this is limited to the discussed one.

Another point that is necessary to make here is that the homogeneity in the semantics of the components within subgroups suggests that the formation of compounds (i.e. their coinage) is driven by analogy. As a result of this, the meaning of a combination appears to be predictable, which supports Štekauer's (2005: 72) idea that the reading a compound receives is not haphazard, but is determined by a number of factors of linguistic and extralinguistic nature. These factors include linguistic characteristics that a word possesses, as well as our knowledge about the actual objects that are denoted by the word. We have seen in *animal + N* and *N + animal* compounds that the latter factor is more important, since the linguistic information provides the general frame, whereas pragmatic knowledge leads us directly to the meaning of a compounded structure.

Can the suggestions mentioned above be extended to other compounds? The concept ANIMAL can be referred to prototypical universal and central concepts of human conceptosphere, which belongs to a more basic domain because it names a concrete object that relates to our senses than nouns denoting abstract entities (Boldyrev 2000). How will the situation change (if at all) if we analyse compounds whose constituents denote more complex concepts?

4.1.5 community + N, N + community compounds: Miscellaneous notes

The constituent word family with the element *community* in my corpus contains 57 items in which the word is used in the modifier position and 64 items in which it is used as a head.

According to the dictionary definition (Oxford online 2012) *community* is understood as a group of individuals, plants or animals that live in the same place or have a particular characteristic in common. The semantics of the noun gives some idea as to what kind of concepts it can be combined with and as to what valence allows for such a combination. A closer look at the semantic content of the noun makes it possible to pinpoint the implied propositions: a community consists of members; these members live in a certain area/place. Therefore, it is possible to predict that at least two valences are active and this should be reflected in conceptual combinations: who/what does the community consist of? (what kind of people, animals, birds, plants), where is the community located? (city, country, village, sea). It can be predicted that the nouns denoting the concepts that are semantically related to these spheres are more likely to be used in N+N combinations with the element *community*.

4.1.6 N + community compounds

According to the results obtained in the course of the data analysis the *N + community* word family receives the reading in which the COMPOSITION relation is realised in 73.4% of the cases, which coincides with the prediction above. This means that the corresponding valence has the highest probability rate (PR 1) in combinations with the noun *community* used as the head of the N+N combination. The next most frequent relation realised in this group is the LOCATION relation with 21.9% of the cases (the corresponding valence has PR 2), which is also consistent with our prediction based on the propositions implied in the semantics of the noun *community*. 31 out of 47 compounds that realise the COMPOSITION relation and 10 out of 14 compounds realizing the LOCATION relation have a frequency of occurrence in the BNC of ≤ 10. So, a relatively large number of compounds together with a low frequency of occurrence in these groups make it evident that these two relations are productively used in the formation of *N + community* compounds, thus pointing to the high PR of the corresponding valences. Two other relations, i.e. PURPOSE, as in *leisure community*, and ESSIVE, as in *model community*, do not seem to be statistically significant and productive (PR 3), however, the realisation of these relations can be an argument for the idea that the number of valences a noun possesses is not limited only to the salient and predictable from the semantic content of a noun concept.

4.1.7 The use of metonymy in *N + community* compounds

Looking at the group of *N + community* compounds in which the COMPOSITION relation is realised, it can be noticed that in 20 out of 47 cases the referent of the modifier explicitly names the member of the community, e.g. *plant community, user community, settler community*, etc. In other cases the link to the member that comprises the community is not so obvious and these cases can be considered metonymical. For example, in the compound *business community*, the modifier *business* refers to a person's occupation or trade or to a commercial activity (Oxford online 2012). The strong association with human activity that is part of the concept BUSINESS makes it possible to understand the compound *business community* as 'the community that consists of businessmen', thus unfolding the compressed structure. In this case we are dealing with some kind of a gap on the level of structural representation, which does not seem to influence the semantics of the whole and does not hamper with the meaning, and, therefore, understanding of this compound. The same is true for other sequences in which the first element does not point directly to the members of the community, e.g. *research community* – 'community of researchers', *mining community* – 'community of miners and people connected with mining in

some way', *whaling community* – 'community of people who are involved in whaling', *speech community* – 'community of people speaking the same language'.

It is important to notice that the referents of the modifiers in such sequences are associated with people, as opposed to animals, plants or birds and the reference to people utilises different routes, e.g. occupation (*retirement community, business community*), language (*speech community, discourse community*), location (*coalfield community, school community*), etc. The semantic packaging of the information implied in the modifiers also seems to differ and in some cases access to it requires more complex elaboration of the meaning on the cognitive level. For example, consider the compounds *hunting community* and *mining community*. In the first case, the modifier is connected with people that hunt, whereas in the second case it includes people that are somehow related to mining as well as the family members of miners. A possible explanation could be that verb-based noun derivatives profile a specified group that includes agents performing the action denoted by the derivational base. This can be seen in the compound *travelling community*. If we compare the latter case with compounds like *whaling community*, *farming community, fishing community*, in which the derivational base of the complex modifier noun is listed as a noun in the first sense, we can see that referents of the modifiers are not limited to the people directly involved into whaling, fishing, mining, etc. Despite the fact that this connection between the morphological status of the derivational base of the modifier and the semantics of its referent may seem quite appealing, it does not work if we come across cases like *engineering community*. Here the derivational base of the modifier is listed as a noun,[93] but the referent does not include people associated with this field. Another explanation can be related to the use of pragmatic knowledge, which allows for a plausible interpretation.[94] However, it is not clear how naïve native speakers process metonymical compounds of this type and, as to my knowledge; there is no research that looks into this issue.

So, the process of creating metonymical compounds in this group, i.e. those whose semantic content includes implied rather than explicit referents, operates on the basis of the modifier rather than the head. Thus, in *N + community* compounds the semantic content of the modifier definitely outweighs the content of the head in terms of the amount of information packed into it. It is also important to note that the analysis of such cases as examples of lexical compression is an overly simplistic approach. The collected corpus contains a considerable number of compounds in

[93] The dictionary entry for the word engineer also contains the information on its use as a verb. However, the etymology of the word suggests that the verbal use developed from the nominal one (Oxford online).
[94] The scope of this research does not allow a further elaboration of this issue; however, it seems to be an interesting question to investigate.

which the referent of the modifier is connected to its denotatum indirectly often naming only one of the significant, typical or associated features.

4.1.8 community + N

It has been shown above that the semantics of the word *community* used in the head position allows us to predict the activity of two valences that are realised as relations in conceptual N+N combinations. However, the word *community* is used in the modifier position quite productively, and now I am going to look into the question of how the semantic content of the noun is revealed in *community + N* constituent word family.

The relational information in *community + N* family is less homogeneous and the compounds of this group receive a wider variety of readings. Based on the statistical analysis of the semantic relations realised in this group, the valence corresponding to the PURPOSE relation (as in *community programme, community staff*, etc.) is the most salient one (PR 1), with 57.9% of all occurrences in this set. The valences realised as the POSSESSION relation (as in *community power, community structure*) and as the TOPIC relation (as in *community law, community research*) are classified as PR 2 valences with 21.1% and 10.5% of occurrences in the analysed data set respectively. The less active valences classified as PR 3 are shown in the compounds that realise the LOCATION relation (*community life*), the INSTRUMENT relation (*community language*) and the SOURCE relation (*community arts*).

Although the range of activated valences realised in *community + N* compounds is quite big, they are not easily predictable from the general semantic content of the noun (based on the dictionary definition). The information we obtain from the various conceptual combinations in the constituent word families reflects our common knowledge about the concept COMMUNITY but this information is not core to the concept and does not seem to be salient and obligatory. The analysis of the ways in which concepts are combined can demonstrate the realisation of the background knowledge that can be seen on the level of language representation. The background linguistic knowledge involved in this process is responsible for the possibility of putting two concepts together, i.e. finding schemas that match (Ryder 1994: 77-79).[95] This implies the use of analogy based on inner statistical information about the grammatical and semantic plausibility of certain N+N formations as well as the similarity in the semantic relations. As a result we can see a certain degree of homogeneity in the semantics of the head elements within relational groups.

[95] See Section 2.6.4.5 for discussion.

The background pragmatic knowledge is based on our understanding of the connections of the objects and phenomena and their interactions in extralinguistic reality. The influence of the pragmatics surfaces in the combinations which are not predictable from the core content of a given concept. Thus the analysis of this kind of interaction allows for building up a much fuller picture of important features associated with the concept denoted by the word, in our case *community*. The same was noticed in the analysis of *animal + N* constituent word family. The impossibility of predicting which valences can be most realised in the modifier concept can be accounted for if we assume that in the process of compound formation and the partial loss of the morphological status by the modifier, it also loses its core semantics. However, the formal shape of the word may be strongly connected with characteristics associated with it on the level of mental representation. These associations are not part of the main meaning of a word. For example, if we think of compounds like *coal community* or *chapel community*, we understand these via the connection with the associated features *coal – mining – miners*, *chapel – congregation – people that come to the church*, etc. The possibility of skipping the literal meaning of the modifier in order to access the associated meaning supports the idea that the modifier partly loses its core meaning.

Based on the information above, we can conclude that the element *community* is productively used in the formation of nominal compounds, which is connected with the complexity of the concept denoted by the word. This complexity is also revealed in the variety of relational readings realised in *community + N* and *N + community* compounds, which supports the idea that the number of active valences for the noun *community* is also large.

4.1.9 business + N, N + business compounds: Miscellaneous notes

The collected corpus contains 168 examples of compounds that contain the element *business*. In 75 cases the element *business* is used as the modifier element; in 93 cases it is the head of the N+N construction, which suggests the preference for the head position. Almost a half of sequences in the head CWF have the frequency ≤10. This is considered to be an indicator of the productive use of the element *business* as the head.

The Oxford dictionary (Oxford online 2012) lists 2 main meanings of the word *business*: (1) 'a person's regular occupation, profession or trade' (this also includes an activity that someone is engaged in, a person's concern, and work that has to be done or matters that need to be attended to), (2) 'a commercial activity' (this also includes profitability of a trade and understanding of business as a commercial house or firm). It should be noted, that although the OED (OED online

2012) distinguishes between these two meanings, the border between them seems fuzzy and the examples of use provided in the dictionary overlap sometimes. The sentences below are taken from the OED (OED online 2012), where they are listed as examples of different meanings the word *business* can realise.

> *For six months, he attended to **farm business**, only playing rugby for Scotland*. (work that has to be done or matters that need to be attended to)
>
> *Now ATS employs more than 110 staff, of which about half are engaged in the **retail business***. (business as a commercial house or firm)

4.1.10 business + N, N + business compounds: Predictability of the relational information

Bearing the ambiguity of the word *business* in mind, we can still make speculations about the propositional structure implied by the meaning of the word, which should be related to the following.

1. what kind of business is it? (*investment business, banking business*); who is involved in business? (*auctinoneer's business, post office's business*);[96] what kind of things/matters need to be done/to be attended to? (*farm business, state business*).
2. what is viewed as a commercial activity? (*marketing business, conference business*); what kind of an activity is a commercial house involved in? (*brewery business, holiday business*).

World knowledge also allows for further speculations about the location of the business (city, village), about the products it produces (food, clothes), about the special instruments or technologies involved (phone, internet), etc.

The examples in the collected corpus demonstrate that the element *business*, when used as the head of compounds, realises different meanings: (1) a commercial activity, e.g. *marketing business*; (2) work that has to be done or matters to be attended to, e.g. *government business*; (3) commercial house or firm, e.g. *brewery business* (OED online). Overall, the compounds that employ the element *business* as a head (especially in its first meaning) demonstrate a high tendency to realising the ESSIVE relation (N1 IS N2), especially with deverbal nouns in the modifier position (*marketing business, banking business, consultancy business*). The frequency of the ESSIVE relation is not surprising since the noun *business* in the meaning 'commercial activity' can be applied to any sphere that can be used as a source of income. However, the nature of this relation

[96] The first elements in these are in the genitive case and are overtly distinct from the forms usually dealt with in this thesis.

is felt to be different and the compounds possess a certain degree of vagueness. This vagueness in meaning seems to be connected with the different meanings the noun *business* in the head position realises. For example, in the compounds like *investment business, entertainment business, equipment business, employment business, insurance business* the ESSIVE relation also has the shade of meaning that can be described as N2 DOES N1. The modifiers in this group are usually deverbal nouns and this contributes to the 'active' meaning of the ESSIVE relation. This makes it possible to suggest that the verbal nature of the modifier is responsible for the head's realising this relation.

The meaning 'commercial house or firm' is realised in compounds like *brewery business*. In this case we can see the connection between the meaning of the head and the meaning of the modifier. The referent of the first element is a place 'where beer is made commercially' (OED online). The meaning of the noun *brewery* contains the ACTIVITY seme but probably because it is a second-step derivation from *brewer* (OED online), rather than from *brew*. The ACTIVITY seme is not so strong, and the priority belongs to the LOCATION seme, which is also part of the semantic information that the noun *business* possesses, i.e. 'commercial house or firm'. In this case we should have an additional locative meaning, but it does not seem to be exactly right, since the interpretation of the compound *brewery business* will involve background information of what the noun *brewery* stands for, i.e. 'business that produces beer'. So, unpacking of the information requires more elaborate analysis in this case.

The meaning of the word *business* that is described as 'work that has to be done or matters to be attended to' is realised in the compound *government business*. The semantic relation changes accordingly, i.e. N1 HAS N2. Although the word *government* is a deverbal noun, it cannot be considered an action nominalisation. The action meaning of the noun *government* is not obvious, as it is more often used in the meaning 'a group of people with the authority to govern a country or a state' (OED online). Since the modifier has a non-verbal referent, the point of overlap of meanings of the modifier and the head is shifted, which results in the different meaning of the link between the components, i.e. 'government HAS business'.

4.1.11 N + business compounds: NDVCs?

One of the groups of the compounds with the element *business* in the head position contains some examples of items in which the first element has a homonymous verb (*transport*, *service*).

transport business, service business, design business

travel business, show business, telephone business, phone business, mail-order business

The terms 'conversion' or 'zero derivation' are avoided deliberately, since both terms imply the necessity of defining the derivational base and direction of conversion, both of these questions being highly arguable if answerable at all. For example, one of the positions is that nouns formed on the basis of verbs by means of conversion are prone to containing the meaning of result or product of the action, but as Lieber (2010: 134) justly points out, the vast majority of these are instead eventive in nature and lack this meaning. So, the semantic structure of conversion items may be the same and is not of much help in solving this 'chicken and egg dilemma'. Therefore, I accept the opinion expressed in Bauer and Valera (2005: 11) that "… the question of directionality is currently unresolved and irresolvable", and will not worry about the conversion from a verb to a noun or conversion from a noun to a verb. The verbal nature of these items and how it contributes to the semantics of compounds is more important, and I will concentrate on that.

Lieber (2010) talks about constructions that she calls *non-affixal (de)verbal compounds* (NDVCs), e.g. *kick ball*, *dog attack*.[97] Compounds of this type have not been classed as a separate category before and used to be viewed as endocentric root compounds. Lieber (2010: 130) provides a comparative/contrastive analysis of NDVCs and synthetic compounds, e.g. *tin-opener*, *burrito assembler*, *city employee* and lists some features typical of this type of compounds.

(1) NDVCs have an interpretation in which one of the elements is construed as an argument of the other, e.g. *dog attack* – 'a dog attacks', *kick ball* – 'to kick a ball';
(2) One of the elements is derived from a verb but there is no overt derivational suffix;
(3) NDVCs are endocentric, i.e. the compound is a hyponym of its head element;
(4) It is not clear whether the verbal element is purely verbal or whether it is represented by a noun formed by conversion from a verb.

Looking at the examples from the collected corpus, it can be seen that three of the cases (*transport business, service business, design business*) seem to satisfy all the criteria, whereas the other two (*travel business, show business*) fail the first one, since the interpretation of *travel*

[97] Lieber (2010: 130) divides NDVCs into two groups: (a) compounds in which the first element is a noun and the second element is formed by conversion from a verb, e.g. dog attack, bee sting; and (b) compounds in which the second element is a simple noun, but the first element might be analysed either as a verb, or as a noun formed by conversion from a verb. Apparently the morphological status of the first element in group (b) is not important for the purposes of the discussion, however, it might be an interesting issue to look into.

business is not 'a business travels' or 'to travel a business' and the interpretation for *show business* is not 'business shows' or 'to show a business'. In the latter case it can be argued that the compound *show business* is lexicalised and, as a result of this, semantically opaque, so the morphological and syntactic features of the constituents are irrelevant for extracting the meaning. But even admitting that, it is not clear why the compound *travel business* does not demonstrate the same pattern. Therefore, it is possible to suggest that although formally the compounds *travel business* and *show business* can be considered NDVCs, their interpretation implies that they are in fact root compounds. The same is true for the examples *telephone business/phone business, mail-order business*.

One feature that Lieber (2010) does not list as one of the characteristics of NDVCs as opposed to synthetic compounds is the implicit semantic relation that they possess. As mentioned earlier, the majority of *N + business* compounds realise the IS relation. However, the meaning of the compounds in this sub-group also implies that 'N2 is a source of N1'. The fact that some of these compounds are NDVCs and some are not seems to be of minor importance here, as on the level of conceptual representation all of the discussed units appear to be the same.

The other big group contains compounds in which the first element does not seem to be derived from a verb. The relational information realised in this group is not very homogeneous. In the following group the head *business* is used in the meaning 'commercial house or firm' and the semantics of the relations can be pinpointed as N2 SOURCE N1, e.g. *holiday business, computer business, property business, food business,* etc.

The essive relation is realised in the following set of compounds, where the first element is non-verbal, e.g. *music business, commodity business, hotel business,* etc. The first element of the compounds in this group denotes an entity that is not understood as a basis of a business *per se*, and interpreting these requires more elaborate processing of the information. These should involve developing a scenario that would describe the way in which the entity denoted by the modifier can be blended (in the conceptual sense) with the head concept *business*. Developing such a scenario involves a necessity to refer to the background knowledge about the constituents that we possess. For example, a *music business* is not exactly 'music is a business', but rather 'a number of different processes involved into creating, producing and delivering music to the public' is understood as a business.

4.2 Summary

The discussion above was aimed to demonstrate that by looking at the semantic relations in constituent families, it is possible to see that some valences have a higher probability for being realised as semantic relations when a noun is used as a constituent of an N+N compound. The probability of different valences for being realised as semantic relations is expected to be different for different nouns and seems to be more strongly connected with the semantics of the word when it is used as the head constituent than as the modifier constituent. In the latter case, analogy was shown to be of importance in the formation of N+N compounds. It was also noted that the complexity of the concept named by a noun allows for a better understanding of the concept through the analysis of the N+N combinations in which it occurs. Simpler concepts, i.e. those naming concrete objects that relate to our senses, like ANIMAL, appear to demonstrate a higher degree of homogeneity in the formation of conceptual combinations than more complex concepts, like BUSINESS or COMMUNITY. This homogeneity can be seen in the semantics and morphological (derivational) characteristics of nouns within the relational groups.

5. Productivity patterns in compounding

The purpose of this section is to further discuss the hypothesis that the constituents comprising a compound should be more productive in one position than in the other. In order to do this, I will discuss the productivity patterns observed in the collected CWFs (constituent word families). This discussion is based on the statistical analysis of the head and modifier CWFs.

The second question that will be looked into here concerns the productivity of the relational pattern realised in a head and modifier CWF and whether it is connected with the productivity on the structural level. I will look into examples from each of the groups and will discuss the factors that contribute to the productivity of a given element in the head or modifier position (semantics of the component, morphological productivity of compound constituents, analogy in the semantics, etc.).

The outline for the discussion is as follows.

1. **Introductory comments.** These include the main issues that are discussed in the chapter, a brief recapping of the theoretical framework, and grouping of the analysed constituent families.
2. **Compounds with the absolute preference for one position.**
3. **Compounds with a strong preference for one position.**
4. **Compounds with a low to moderate preference for one position.**
5. **Summary of the findings**.

The discussion of each group looks into the possibility of connection between the productive use of a constituent in an N+N compound and the productive utilisation of one of the semantic relations. In order to do that, the semantics of the constituent is looked into and an attempt is made to draw a parallel between the meaning of the word and the semantic relation predominantly realised in the constituent family.

5.1 Introductory comments

The analysis of the collected constituent word families (CWFs) data presented below aims at accounting for the productivity of the key constituents in the head or modifier position based on the principles of paradigmatic analogy in the formation of compounds, as discussed in Section 2.6. The analysis accounts for the structural patterns noted in different groups of constituent word families and further discusses the hypothesis that there is a connection between the positional productivity of the constituent and the relational reading realised by this constituent in the head or modifier position. It will also be shown that the probability rates for the relational information realised in CWFs are higher for head CWFs than for modifier CWFs, which supports the hypothesis about the leading role of the head, whose properties allow for combinations with some modifiers but not with others. The higher probability rates of relational information in head CWFs are shown to be connected with the semantics of the modifiers, which demonstrate the possibility of being referred to a limited number of semantic fields. The uniformity in the semantics of heads is not so apparent in modifier CWFs.

I will attempt to provide a comparative analysis of some of the most representative modifier and head CWFs that, as it appears, demonstrate different degrees of positional productivity, or lexical specification. It is claimed here that different degrees of positional productivity on the structural level can be seen in the different degrees of preference for being used in one position in the modifier and head constituent word families, i.e. families united by one and the same constituent. An approach that does not make a sharp distinction between abstract schemas and analogy, but rather views these as two extreme points on a scale, allows us to explain why different words demonstrate different degrees of lexical specification. It also allows us to speculate about why different constituents are used in one position with different degrees of productivity.

Issues addressed in this chapter

- The formation of new complex lexemes is based on the paradigms of similar existing complex lexemes and their structural characteristics rather than on abstract rules.
- The form of new compound coinages relies on the formal and semantic properties that the constituent words in the new combinations share with other compounds these constituents occur in.
- There is a possibility of a connection between the positional productivity of the element and the relational reading realised by this element in one position, since the formation of compounds is driven by analogy.

- The paradigmatic approach explains why more than one relation is realised in one CWF.
- Higher probability rates of relational information in head CWFs are connected with the semantics of the modifiers, which demonstrate the possibility of being referred to a limited number of semantic fields.
- The role of the head is that of allowing combinations with some modifiers but not the others.

As the statistical analysis shows, a noun's preference for being used in one position increases with family size. The analysis below looks into constituent families that demonstrate different degrees of preference for one position and attempts at drawing the connection between the constituent family size and the use of semantic relation.

The collected constituent families can be divided into three groups depending on the degree of positional preference: (1) absolute (100% of preference for one position), (2) strong (≥80% of preference for one position) and (3) low to moderate (51%-79% of preference for one position).[98] I will look into ten families with the constituents:

(1) Absolute preference (*future, figure*);
(2) Strong preference for the modifier position (*health, family*) and the head position (*programme, problem*);
(3) Low to moderate preference for the modifier position (*court, wind*) and the head position (*product, disease*).

5.1.1 Compounds with an absolute preference for one position

5.1.1.1 *future + N* compounds: Miscellaneous notes

This CWF demonstrates absolute preference for the constituent being used in the modifier position, which might be connected with the morphology of the word *future*. The dictionaries list the word as a noun and as adjective; with the adjectival status in *future + N* combinations. This can be explained by the fact that *future* was adopted into English (via Old French) from Latin *futures*, future participle of *esse* 'be' and the borrowed form inherited some morphological/functional characteristics of its etymon. Although there is no clear evidence as to which of the uses of the word *future* (adjective or noun) was initially preferable, since both are registered in the late 14th century (OED online), it is often the case that the original morphological properties of the word are preserved in the recipient language (Grzega 2003). Some dictionaries, like Merriam Webster's

[98] Grouping of compounds according to the degree of positional preference is presented in Appendix 6.

dictionary (Webster online 2012), even list the word as an adjective as its primary use. So, it might be the case that the borrowed item started to be used as a noun later.

The collected data shows that *future* occurs only as a modifier of two-word sequences, which might be the main reason for listing the word as an adjective in the modifier position. Thus, functional characteristics can be considered more important than structural peculiarities by the lexicographers, who have to take the most practical decision when compiling a dictionary entry (Connor 2010: 34).

Based on the approach taken in this research, a word is considered to be a noun and is included in the analysis if it has the structural characteristics typical of a noun and can be used in the role of a predicative surfacing as a noun (it behaves as a noun). It is assumed that the word has the structural characteristics typical of a noun, but, on the level of two-word sequences with a noun-head, it fulfils the attributive function (as is the case with the modifiers that are listed as nouns in dictionaries).

The opacity of the morphological status of the word *future* can be considered a clear-cut example of the high degree of lexical specification when the constituent seems to be prescribed to be used in the modifier position. Opacity of the morphological status and the lack of grammatical properties of the word in *future + N* sequences contribute to the development of the constructional idiom, when novel coinages are formed using the analogy that is crystallised to the degree that the use of the word in the rightmost position of an N+N sequence does not seem to be customary. The perception of the word *future* as an adjective strengthened by its use as an attribute accounts for the collocational potential of the word, and contributes to its sole use in the modifier position in the collected corpus.

It is difficult to say whether the lack of *N + future* compounds in the BNC and, consequently, in the collected corpus is accidental or whether it is an interesting fact about English. Adj + N combinations like *conditional future* or *uncertain future* demonstrate that the noun *future* can be submodified, however, it is not clear why there do not seem to be any examples where *future* (in whichever of its meanings) is modified by a noun.

5.1.1.2 *future + N* compounds: Productivity of the semantic relation

The analysis of the semantic relations realised in *future + N* compounds demonstrates a strong correspondence with the semantics of the modifier, with the valence realised as the TIME relation being prevalent (89%) in the CWF. This may mean that the valence corresponding to the TIME relation is most likely to be utilised in the process of formation of *future + N* compounds, which suggests high productivity of the temporal relation. The data support this suggestion; only three of the sequences in the TIME relational group (*future generation(s), future date(s), future year(s)*) have a relatively high (more than 100 tokens) frequency of occurrence in the BNC. The activation of the temporal valence does not seem to be dependent on the semantics of the head in *future + N* combinations, which might be due to the high collocational potential of the word and the possibility of viewing practically every phenomena of extralinguistic reality in a temporal perspective.

5.1.1.3 *N + figure* compounds: Miscellaneous notes

The key word of this CWF does not give rise to speculations as far as its morphological status in N+N combinations is concerned. The word *figure* realises its primary meaning in the collected CWF, i.e. 'a number, especially one which forms part of official statistics or relates to the financial performance of a company', as in *management figures, government figures* (Oxford online 2012).

A closer look at the morphology of the modifier elements in this CWF shows that the modifier is likely (46 out of 69 cases) to be a deverbal noun and to name a process or phenomenon (*government figures, population figures*) or a noun that has a homonymous verb (*shock figure, cost figure, survey figures*). Examples in which other meanings of the word are realised, as in s*he's got a good figure* or *a figure of speech,* were not included in the corpus because these are believed to be homonyms rather than different meanings of the word *figure*, since they have nothing to do with numbers.

The collected data demonstrates that the element *figure* in its primary meaning is used in the head position of N+N combinations in the absolute majority of cases. This means that the element *figure* is more likely to be modified and is highly unlikely to be a modifier. The absolute preference for the head position implies a high degree of lexical specification of the element on the level of the N+N construction. The number of coinages that have a low frequency of occurrence in the BNC is quite high (only two of the compounds, *unemployment figure* and *sales figure*, have a frequency of more than 100), which is an indicator of the productivity of the *N + figure* pattern. However, it is

not right to claim that the *N + figure* pattern can be considered a constructional idiom in the same degree as *future + N* compounds, since the semantic relations realised in the *N + figure* CWF are not so homogeneous.

The question that seems interesting concerns the reasons for some nouns to be predominantly used in the head position. The most obvious answer is the influence of analogy, which drives the formation of new coinages based on existing models. It is worth pointing out that the example of *N + figur*e CWF clearly shows that the analogy does not have to be driven by one (often lexicalised) model word, but rather by the whole paradigm. A number of compounds in the collected data set seem to be coined randomly for immediate purposes or for the purposes of further specification of the head without producing the effect of novelty, despite the fact that more conventional items already exist in the language, e.g. *listenership figures, viewing figures* vs. *audience figures*, *death figures* vs. *mortality figures*, *arrest figures* vs. *police figures*. The influence of analogy can also be seen on the morphological level in the examples where modifiers are deverbal nouns derived by means of certain suffixes, e.g. *establishment figure, unemployment figure, employment figure, settlement figure, requirement figures, recruitment figures, replacement figures; inflation figure, production figure, circulation figures, consumption figures*, etc. In the vast majority of cases the homogeneity in the morphology goes in parallel with the homogeneity in the semantics of the whole compound. However, it is not right to claim that this should always be the case. For example, different relations are realised in the combinations *establishment figure*[99] and *employment figure*, which implies that it is not impossible for structurally similar modifiers to contribute to the overall meaning of the coinage in different ways.

[99] The compound *establishment figure* is ambiguous and can refer to a person or to a number. Compare the examples from the BNC:

*In the 1890s a prominent **establishment figure** , General Tani Kanjó, had called for greater attention to be given to the interests of the farming population*.

*Although the perception that Phase 3 schools were under-resourced was widespread, the effectiveness of the PNP staffing enhancement was certainly not a consequence of numbers alone, any more than it is when schools are staffed at the conventional **establishment figure.***

Only cases in which the latter meaning is realised are included in the analysis. The same was done for other compounds that are ambiguous, e.g. *authority figures, party figures*, etc.

5.1.1.4 *N + figure* compounds: Productivity of the semantic relation

The semantics of the word *figure* allows for the presupposition as to where a given figure comes from, and what the figure is about, thus profiling the valences which should attract the nouns that can serve as a source or the topic of data. So, the semantics of the noun *figure* used as the head predetermines the semantics of the modifier that the former can be aligned with (something that can serve as the topic or the source of the data). The statistical analysis of the CWF confirms this suggestion with PR1 valence realised as the TOPIC relation, e.g. *crime figure, employment figure* (60.5%) and PR2 valence realised as the SOURCE relation *revenue figures* (18.5%). The other relations realised in the *N + figure* CWF include ESSIVE *record figure* (PR2), and CAUSE *shock figures*, LOCATION *headline figure*, COMPOSITION *passenger figures* and PURPOSE *establishment figure* (PR3).

The analysis of the modifiers in this CWF shows what kind of objects and phenomena can be profiled in accordance with the active valences of the head noun. The modifiers in this relational group can be divided into two groups based on the meaning they convey. By far the biggest group (59 out of 69 compounds) includes compounds in which the modifier noun refers to processes or phenomena that relate to human activities, e.g. *employment figure, performance figure, mortality figure*. The understanding of the relational information as the TOPIC relation presupposes the reference to the context in which the processes and phenomena outlined by the modifier can take place. The direct reference to the source of the information, i.e. a company or organisation, is realised in compounds in which the modifier serves as the source of some figures. However, it can be noticed that the subdivision of the compounds in this CWF into TOPIC and SOURCE ones is not so sharp, and often the former are felt to be indirectly connected with the SOURCE reading. For example, in the case of *mortality figure*, the modifier *mortality* can be perceived as the source of the figure, since if there was no mortality, there would be no pragmatic need for a compound that would describe this. It seems to me that the semantics of the head contributes to the elaboration of the content of the modifier concept in this direction.

The other relations realised in this CWF do not appear to be easily predictable based on the semantic content of the head, which supports the argument about the potentially present valences that may or may not be realised as relations.

We have seen that the word *figure* is more productively used with deverbal nouns in the modifier position; the deverbal noun is most likely to name a process of some kind or a phenomenon related to human activity. This preference is attributed to the use of analogy in the formation of

N+N compounds. It is also suggested that analogy on the structural level does not always have to coincide with analogy in the semantics of the coinages, however in the vast majority of examples it is still the case.

A relatively high degree of transparency of the semantics of the compounds in this CWF, and the consistent activation of a limited number of valences in the conceptual representation of the head, can be the indicators of the productivity of the pattern, which is triggered by analogy on the level of the paradigm.

5.2 Compounds with a strong preference for one position.

5.2.1 Preference for the modifier position. Comments on the morphology of modifiers and modifier CWFs of this group

It has been suggested elsewhere that a strong preference for the leftmost position may sometimes be held responsible for the confusion concerning the morphological status of the modifier element of N+N combinations. Some researchers (e.g. Giegerich 2004, 2005) point out the semantics of the modifier element as the factor that is responsible for a word a noun or an adjective when it is used in the attributive position. For example, words denoting a material (*wood, stone, cotton*) used in the modifier position are considered to be similar to adjectives. However, the analysis of the dictionary entries suggests that this does not seem to be the case even for nouns denoting material. It might rather be that the modifiers which are often pinpointed as adjectives when used in the attributive position do not have an adjective derivative that is based on a given noun. For example, the modifiers in this group that are defined as clear nouns, i.e. *air, health, heart, chocolate, art, lemon,* can be used as a base for deriving an adjective, e.g. *airy, healthy, hearty, chocolaty, arty, lemony* respectively.[100] The constituent *family* in its attributive use (*family + N* sequences) is listed as an adjective (Oxford online 2012) and, as predicted, there is no adjective obviously derived from it.[101] The same can be said about the confusing cases of the words *animal* and *future* that were discussed above. Thus, we can suggest that the preference for the modifier position is not connected with the opacity of the word's morphological status. An influencing factor on the 'adjective feel' is the necessity of using one and the same structural form in different functions.

[100] It seems that the morphological unification is not a requirement, neither is the unanimity of derivational meaning.
[101] It can be argued that the form *familial* can be considered a derivative from the noun *family*. However, in this case the orthographical and etymological similarities turn out to be misleading.

There does not seem to be any recurrent morphological patterns that characterise head nouns in this group of modifier CWFs. Semantically, this group demonstrates a rather broad range of relational patterns across all compounds. However, the statistical analysis of each CWF taken separately demonstrates the preference for one or two valences that have a higher probability of being activated (PR1 valences). The active utilisation of one (maximum two) valences can be explained by the peculiarities in the semantics of some of the components. For example, the constituents *lemon* and *chocolate* have a strong tendency to be combined with nouns denoting kinds of food in which chocolate and lemon are used as complementary or base ingredients for certain dishes/foods (*chocolate ice-cream, chocolate drink, chocolate truffle, chocolate flakes, lemon jelly, lemon cake, lemon sorbet, lemon marmalade*). The linguistic analysis demonstrates that the differences in the relational patterns between the compounds in which these constituents are used as the main ingredients (COMPOSITION) and the compounds where they are used as subordinates (POSSESSION) is not in any way reflected in the semantics of the head nouns, but is rather dependent on what we know about how things work in extralinguistic reality.

However, it is necessary to point out that *chocolate + N* and *lemon + N* CWFs are the only ones in the collected data set with such unanimity in the semantics of the head nouns. In general, the semantics of the head nouns is difficult to predict from the modifiers, even within one relational group. For example, in the largest relational group of *art + N* compounds the modifier concept *ART* is profiled as the TOPIC of something denoted by the head. The head nouns in this group can denote a place (*art gallery, art school, art museum*); a field of study and people involved in this study (*art history, art education, art student*); as well as a number of other things that would be difficult to group in terms of their semantics (*art class, art magazine, art show, art news*). In the *air + N* CWF, the prevalent relation is LOCATION, and the modifier *air* in this relational group is profiled as a place. So, we would expect the head noun to name a process, phenomenon or object (*air movement, air crash, air ambulance*) but not nouns referring to quality (*air superiority*), a period (*airtime*) or a military unit (*air force*). These, however, do not seem to cause any difficulty if we analyse the semantics of the heads. The same holds true for the other constituents in this group. The lack of unanimity in the semantics of the heads appears to be connected with the inability to predict the semantics of the rightmost component.

It has been suggested before that the productivity of the semantic relation should be correlated with the semantics of the components within the relational groups, i.e. the similarity in the abstract schemas. This means that the modifier should be compounded with heads that are

characterised by similar semantics. In this case the similarity in the abstract schemas of the heads should suggest similarity in valences that are most likely to be activated. However, it appears that this suggestion works only on the limited basis in the analysed modifier CWFs. It might be that this suggestion is difficult to check for the modifier CWFs, since (as has been shown before) the head, rather than the modifier, is responsible for the formation of N+N compounds, and the head's abstract schema and valence structure predetermine what kind of lexical item it can be compounded with and what relational reading a compound will receive. The role of the modifier is that of overall meaning construal by characterising the head in the way that is most appropriate for the needs of communication.

I will now move on to a more detailed analysis of the most representative modifier CWFs. Two families with the constituents *health* and *family* will be analysed. I will first give a brief overview of the two CWFs and then look at each of them in more detail.

The first word (*health*) is listed as a noun, whereas the element *family* has a dubious status, and is listed as an adjective when used as a modifier in some dictionaries (Oxford online 2012, Webster online 2012), but not the others. For example, the Collins dictionary (Collins online 2012) and Cambridge dictionary (Cambridge online 2011) list it as a noun irrespective of the attributive function.

The CWFs discussed here are all characterised by the strong preference of the constituents (*health* and *family*) for the modifier position, i.e. the constituent is used in the head position in >10% of the cases in the collected corpus.

Dictionaries list a number of compounds with the discussed elements in the modifier position (e.g. *health care, health centre, health certificate; family man, family tree, family circle*, etc.). These compounds, though established, are still relatively semantically transparent in most cases. The number of compounds registered in dictionaries in which these elements occur in the head position is limited (e.g. *product family*). Although, as suggested by Booij (2005), lexicalised items may trigger the formation of structurally similar items, it seems necessary for a considerably large number of items that form a paradigm to be lexicalised in order to encourage further coining of structurally similar items (Krott 2009, Krott et al. 2007). The question of how large a model paradigm should be in order to create analogical formations is yet unclear. Bauer (1983: 51), for example, believes that word-formation by means of analogy should be opposed to word-formation by means of rules, since in the case of analogy a new formation is clearly based on one already existing lexeme and this process is not productive. Booij (2005) claims that it is often

impossible to have a clear-cut-distinction between rules and analogy, but it is possible for one lexicalised item to trigger the process of analogical formation. Since lexicalised compounds are not looked into in this research, it is difficult for me to make any comments on this issue.

5.2.1.1 *family + N* compounds: Miscellaneous notes

The dictionaries (Collins online 2012, Oxford online 2012, Webster online 2012) list the word *family* as an adjective when it is used in the attributive role. As was suggested elsewhere, this can be related to the lack of the apparent adjectival derivative which uses the noun *family* as a derivational base. However, the counter argument in this case can be the existence of the adjective *familial*, whose orthographic shape implies the relation between the words. It is clear that the adjective *familial* is etymologically related to the word *family*, and based on information in the Oxford dictionary (Oxford online 2012) both words are loans from Latin (via French). However, the noun *family* was borrowed at a much earlier period (the beginning of the 15th century) to refer to servants in the household and developed its modern sense, 'those related by blood', in the mid 17th century. The noun is suggested to have developed its adjectival meaning, i.e. 'suitable for all family members' in the early 19th century. The considerable semantic changes in the meaning of the loanword *family* mean that it has become semantically assimilated in the language and has acquired the life of its own. The adjective *familial* has a limited sphere of usage (mostly medical contexts and contexts relating to natural science), has a relatively low frequency of occurrence (at least in the BNC), and at the current moment it does not motivate the formation of derivatives,[102] which signifies its low degree of assimilation. The latter is usually characterised by low frequency, limited collocational potential and the specialised area of usage of the adopted form (Volodarskaya 2002: 109). All of these characterise the loan and may be attributed to *familial* being borrowed at a much later period (early 20th century) than the noun *family*, which does not have such restrictions. So, the pair *family* and *familial* can be considered good examples of the differences in the degree of assimilation of independent loans whose etymons are derivationally related in the source language. Based on the collected compound data, it can be predicted that the adjective *familial* is unlikely to become widely used since the noun *family* seems to fulfil the attributive functions successfully realising the same sense as *familial*.

The example of *family vs. familial* provides some insights into the question of competition between a noun and an adjective derived from a noun (or related to a noun) in an N+N

[102] Although the noun *family* does not seem to motivate the formation of affixal derivatives, it is productively used in the formation of compounds, whereas *familial* does not motivate the formation of derivatives. The fact that the word familial has a very limited sphere of usage also implies a low collocational potential.

construction. We have seen that the choice of an adjective or a noun may depend on the time of the adoption of the adjective and the noun. The other factors that should be named are: the differences in the semantics between the noun in the attributive role and an adjective (*beauty queen ≠ beautiful queen*), the existence of the object of the extralinguistic reality that can be denoted by one construction but not the other, i.e. we cannot use the compound *nocturnal owl* to refer to a person that goes to bed late, which is the meaning of the compound *night owl*; neither can we say that a *night owl* can be used to denote any of the Strigiformes (owls), since a *night owl* is not an owl at all. We can also add the stylistic factor. For example, in the cases like *heart disease* vs. *cardiac disease*, *spring flower* vs. *vernal flower,* the stylistic factor predetermines the use of the adjectival modifier in the formal context and the use of the noun in neutral contexts.[103]

The productive use of *family + N* schema (rather than *N + family* schema) is reflected in the statistics, with >10% of compounds employing the word *family* as a head.

5.2.1.2 *family + N* compounds: Productivity of the semantic relation

We discussed above that the word *family* in the attributive position is listed in dictionaries as an adjective and realises the meaning 'suitable for all family members'. Based on the collected data, this meaning is not very frequent in *family + N* CWF and is found only in the sequences *family film* and *family doctor.*[104] In all other 71 *family + N* compounds the word *family* seems to project a more general meaning 'relating to a family in a certain way'. The way in which *family* is connected with the head noun does not appear to be predictable from the semantics of the component, but is rather determined by pragmatic or general knowledge. Based on the dictionary definition, a prototypical family is 'a group of people consisting of two people and their children living together' (Oxford online 2012), however, in the modern world, the prototype seems to be changing. FAMILY is a very complex concept, and this complexity arises from the fact that this is one of the basic concepts that absolutely any person can relate to and therefore it is loaded with information that has roots in our emotions, experience (positive or negative), memories, etc. We will not be discussing the structure of the concept FAMILY in detail here, and will move on to the discussion of the productivity of the semantic relations.

[103] However, there are cases where this doesn't work, e.g. *atom bomb* vs. *atomic bomb*. The competition between an adjective and noun is an interesting issue and is worth further investigation. This is one of the topics that I would like to pursue in my further work.

[104] It is interesting that the compound *family movie* does not occur in the BNC and the compound *family film(s)* only occurs five times. Absence of *family movie* may be explained by the fact that a British corpus was used and because of the dates of the corpus (1980s to early 1990s). Probably, *movies* had not fully caught on as a term in Britain.

In the case of *family + N* compounds, we can see that the most productive relation is that of POSSESSION (*family money, family health, family tradition, family house*) with 41.7% of occurrences, followed by PURPOSE (*family doctor, family trust, family room*) with 31.9%, both of which should be considered realisations of PR1 valences. Based on the semantic content, as listed in dictionaries, the POSSESSION relation does not appear to be in direct correlation with any of the meanings of the word. At the same time, this relation is realised in the collected *family + N* CWF consistently, which means that the corresponding valence should be highly predictable from the semantics of the word. However, this does not seem to be the case and understanding of *family* as 'a possessor' is tied to our personal knowledge and experience rather than to the meaning of the word. The same seems to be true for the compounds realising the LOCATION relation (PR2), e.g. *family atmosphere, family relations*. Understanding/ perceiving the concept FAMILY as a kind of space (mostly in some abstract sense) demonstrates how strong associative ties between the concepts FAMILY and HOME are utilised and reflected in the meaning of nominal compounds. The high probability rate of the valences realised as the POSSESSION and the PURPOSE relations can be attributed to the influence of analogy. The relatively large number of compounds in this CWF realising the relations which are not predictable from the semantic content of the element *family* promotes the corresponding valences, and makes them the most likely candidates for activation when new *family + N* compounds are coined.

The valences realised as the TOPIC relation (*family law, family history*), the COMPOSITION relation (*family member*) and the INSTRUMENT relation (*family system*) are assigned PR3 status. The fact that not all of the relations can be predicted from the semantic content of the word *family* when used in the modifier position supports our suggestion that a noun concept can realise any relational information in conceptual combinations because of the valences potentially present in the conceptual structure of the component. The relations realised in N+N sequences are difficult to predict for the modifier concept, since, as is assumed here, it is the construction schema of the head concept that is primarily responsible for assigning the reading that an N+N compound receives.

It is also worth noticing that the concept FAMILY can be referred to one of the scenario concepts discussed above. One of the characteristics of such concepts is a high collocational potential, which can be seen in the combinations they occur in and a variety of meanings that these combinations convey. The meanings that are profiled in language constructions employing such concepts are not explicitly present in the semantics of the concept but are stored on the

conceptual level. In linguistic terms such meanings can be called secondary predicates/arguments and their activation cannot be predicted easily.

5.2.1.3 *health + N* compounds: Miscellaneous notes

The other CWF that will be analysed in this group is comprised of compounds that contain the element *health* in the modifier position. The collected corpus contains 56 *health + N* compounds as opposed to three compounds in which *health* is used as the head (*child health, family health* and *animal health*). The notion of health is one of the vital notions in our life and there are a number of arguments concerning the issue of defining health. The famous New Zealand writer, Katherine Mansfield, whose definition is often cited, wrote: "By health I mean the power to live a full, adult, living, breathing life in close contact with what I love … I want to be all that I am capable of becoming…". The WHO constitution defines health as 'a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity'. The Oxford dictionary (Oxford online 2012) provides quite simplistic definitions of health: 'the state of being free from illness or injury', 'a person's mental or physical condition'. The definition given in the Cambridge dictionary (Cambridge online 2011) is a bit more extended but still does not reflect the complexity of the concept: 'the condition of the body and the degree to which it is free from illness, or the state of being well'. Although both dictionary definitions are based on negation, which is often considered bad practice, it is justified by the fact that we do not think about health issues if the lack of health does not cause us problems. The important information that we can extract from the definitions concerning its appearance in N+N constructions is that some (animated) object can be characterised as healthy or not (has health or not). However, the possessor of health should take the modifier position in an N+N sequence (and this can be seen in the three *N + health* compounds in our corpus, e.g. child HAS health, animal HAS health, family HAS health), which is not very helpful for making predictions concerning the semantics of heads and the semantic relations that can be realised in *health + N* constructions.

The noun *health* motivates the formation of the adjective *healthy*, which implies the possession of health. Therefore, *healthy + N* combinations convey the meaning different from those of *health + N* compounds. In the former case, the meaning is limited to that of 'N possesses health' or 'N is classified as being healthy', whereas in the latter case, at some level of mental representation, the possibility of the absence of health may be implied. This seems to be the case with the absolute majority of *health + N* compounds in the collected CWF. The modifier *health* can be replaced with *healthy* and still more or less preserve the meaning only in one compound in our corpus, i.e.

health food. But even there we can feel that the modifier *health* in this compound bears some implication connected to the lack of health.

5.2.1.4 *health + N* compounds: Productivity of the semantic relation

It has been shown above that the semantic content of the noun *health* is not of much help for describing the combinatorial properties of the word when it is used in the attributive role of N+N sequences.

The biggest relational group in this CWF is the PURPOSE group (64.3% of the cases), in which the modifier *health* is profiled as an aim of something. However, the profiling appears to be different and this group can be roughly divided into two subgroups depending on the meaning conveyed by the modifier. The heads of the compounds comprising the first subgroup are intended for promoting health as a condition or state of physical/mental wellbeing of an individual as well as a group of people, e.g. *health medicine, health food, health club, health facilities*. In the second subgroup the modifier *health* is understood as a societal construct, therefore, health is perceived as a value for a group of people (society, community, nation, etc.) rather than an individual's wellbeing e.g. *health sector, health system, health union, health council*. In this case health is profiled as a basic human right for which the society, state, government, etc. hold responsibility. In the first group we can see the direct reference to the meaning but it becomes less explicit in the second one, and this is believed to be induced by the semantics of the head nouns in the second group, that are generally used in the reference to other things that require monitoring, supervising, promotion, etc. This relation is considered to be productive for *health + N* compounds, despite the fact that the valence corresponding to the PURPOSE relation does not seem so obvious for the concept HEALTH.[105] So, we can see that the valence realised as the PURPOSE relation can be considered PR1 valence for the noun *health* in the modifier position.

The second most common relation realised in this group is the TOPIC relation (with 17.9% of occurrences) and, based on the statistical analysis, the corresponding valence is assigned PR2 status, e.g. *health act, health policy, health research, health statistics, health experts*. The meaning conveyed by the modifier is felt to be related to a group of people, especially those whose health is not ideal. The topical relation is directed at the modifier, thus profiling health as the topic of something described by the head noun.

[105] The collected corpus does not contain enough data that would allow for a valid comparison between *health + N* and *N + health* compounds, so it is difficult to make speculations concerning the activity of the PURPOSE valence when the element *health* is used in the head position.

The other relations realised in this group include CAUSE (*health problems*), LOCATION (*health farm*) and POSSESSION (*health status*). The number of compounds realising these relations is marginal but the fact of their realisations points to the valences that are potentially present in the conceptual structure of the word.

5.3 Preference for the head position

5.3.1 Introductory comments

According to the idea investigated in this chapter, it is the semantic content of the head (not the modifier) that is responsible for the relational reading a compound receives. Therefore it is possible to predict the relational information realised in compounds based on the semantics of head nouns. To determine whether this proposal does or does not have merit, I will be looking at dictionary definitions that are assumed to be helpful in defining the possible valence structure and propositions implied in the semantics of a noun. The predicted propositions will be checked against the data in the collected head CWFs containing the elements *programme* and *problem*.

5.3.1.1 *N + programme* compounds: Miscellaneous notes

The constituent *programme* is used in 111 compounds, 105 of which utilise the noun as the head of an N+N sequence.[106] The overwhelming preference for the head position implies that there might be a constructional scheme which encourages the formation of *N + programme* compounds, but not *programme + N* compounds.

The word *programme* is listed in the dictionaries (Collins online 2012, Oxford online 2012, Webster online 2012) as a noun and as a verb, which makes *N + programme* compounds ambiguous in terms of their morphology when they are taken in isolation. The head element can have a homonymous verb in examples like *computer programme*, since the interpretation 'to $(N2)_V$ N1' is possible, i.e. 'to programme a computer'. Ambiguous cases like that were avoided where possible; however, a certain amount of ambiguity remains in all *N + programme* compounds, since it is possible to say 'to programme a course' or 'to programme funding' and get *a course programme* and *a funding programme* as a result.

The meaning realised by the constituent in the collected *N + programme* CWF is that of 'a set of related measures or activities with a particular long-term aim' (Oxford online 2012), e.g. *education programme, study programme, training programme*. Compounds in which the element

[106] The alternative spelling, i.e. *program*, was also checked and the examples were included in the collected data set.

programme realises other meanings, e.g. *television programme*, were excluded due to the limitations on the search outlined in Section 3.2.3. However, in some cases the meaning of the constituent is ambiguous. For example, in *youth programme*, the head noun can be understood as 'an item broadcasted on radio or television'; in *seminar programme*, the meaning of the head can be that of 'a sheet or a booklet giving details of items or performers at an event or performance'. In the cases where such ambiguity was implied, the mini-contexts in which the ambiguous compound occurs were checked.

5.3.1.2 *N + programme* compounds: Productivity of the semantic relation

Based on the dictionary definition, i.e. 'a set of related measures or activities with a particular long-term aim' (Oxford online 2012), we can suggest that the most active valence (PR1) for the element *programme* should be the one connected with the aim or expected result of some programme, which should, in its turn, be realised as the PURPOSE relation. For example, the ultimate purpose of an *education programme* is for someone to be educated, a *funding programme* is aimed at providing funds to something or someone, a *discussion programme* presupposes that some matters on the agenda will be eventually discussed (unless it is a programme on the radio that involves discussion), etc. Our pragmatic knowledge also suggests that, since a programme is a man-made thing which is designed for a certain period of time, the valences corresponding to the relations of SOURCE and TIME can be also active. However, these valences cannot be equally active, because the source of information projected in the case of their activation is our general knowledge about the way things work in reality.

The statistical analysis of the relational link in the *N + programme* CWF supports our prediction, with the PR1 valence realised as the PURPOSE relation in 94.3% of all cases. In the case of the noun *programme* we can see how the most important valence is strongly implied by the semantic content of the noun. The general sense of 'a definite plan or scheme' realised by the head noun is suggestive of the intended result or outcome outlined by the modifier. The intentional PURPOSE meaning is projected by the head *programme* onto the modifier concept, which is also reflected in the directionality of the relation, i.e. N2 FOR N1. Interestingly, the analysed dictionaries (Cambridge online 2011, Collins online 2012, Longman online 2012, Oxford online 2012, Webster online 2012) include the reference to the goal/aim/purpose as part of the semantic content of the word. This can be considered an example of how the propositional content (which is expected to be stored on the deeper level of conceptual representation) surfaces on the semantic level. The consistent realisation of the PURPOSE relational pattern can also be attributed to this.

It was noted before that modifiers within relational groups of head CWFs may share some semantic and sometimes morphological properties, as the abstract schema of the head concept is definitive for deciding what kind of modifier concepts it can be combined with and what the modifier's linguistic (as well as extralinguistic) characteristics should be. This, together with analogy involved in the production of compounds with somewhat similar characteristics, should surface on the level of linguistic representation. In the collected *N + programme* CWF, the modifier is expressed by a deverbal noun (*investment programme, environment programme, training programme*) or a noun that has a homonymous verb (*work programme, research programme, control programme*) in 68 out of 98 compounds in this relational group. The fact that the modifier is strongly associated with a verb in this case contributes to the implied meaning of the intended purpose or result, which is projected by the semantics of the head noun *programme*. Although the modifiers of the other compounds in this relational group do not seem to demonstrate such a strong unity of meaning, they are generally projected as a goal, purpose or result of the head noun, e.g. *literacy programme* (someone is literate as a result), *safety programme* (safety is provided as a result), *power programme* (delivering/distributing power is the aim of the programme), *fitness programme* (is intended for becoming fit), *certificate programme* (gaining a certificate is the purpose of the programme).

The *N + programme* CWF also demonstrates activation of other valences, all of which are considered PR3 valences (based on their occurrence in the collected corpus) and are realised as the following relations: NEG PURPOSE (*weapons programme*), TIME (*winter programme*), ESSIVE (*core programme*), SOURCE (*government programme*). It was suggested above that profiling of the concept PROGRAMME in such a way is not directly related to the semantic content, but is connected with other kinds of knowledge. Activation of these valences is not at all frequent as shown by the data in the corpus, which allows for a suggestion that the realisation of these relations is not very productive for *N + programme* compounds.

5.3.1.3 *N + problem* compounds: Miscellaneous notes

The collected *N + problem* CWF contains 117 compounds, as opposed to 10 compounds in which the constituent *problem* is used in the modifier position. This signifies the productivity of the constituent *problem* as a head rather than as a modifier.

The word *problem* is defined as 'a matter or situation that is unwelcome or harmful and needing to be dealt with and overcome' (Oxford online 2012), 'a situation that causes difficulties' (Longman online 2012), 'an intricate, unsettled question; a source of perplexity, distress, vexation'

(Webster online 2012), 'something or someone that is difficult to deal with' (Collins online 2012). As we can see from the definitions, the word *problem* has strong causative semantics and negative connotations, which are expected to influence the semantics of N+N compounds.

Interestingly, some of the dictionaries (Collins online 2012, Webster online 2012) also list the word both as a noun and as an adjective when it is used in the attributive role, e.g. *problem child*. Although we are not concentrating on the *problem + N* CWF, the reasons for the decision about the adjectival status of *problem* as a modifier are somewhat obscure, since the word cannot be used in the predicative position without an article (**This child is problem*). Even if we assume that *problem* belongs to the class of non-predicate adjectives, e.g. *bovine, dental, viral*, these still have to have homonymous nouns that we may presume have been related to,[107] which is considered to be one of the characteristic features of this class.

5.3.1.4 *N + problem* compounds: Productivity of the semantic relation

The dictionary definitions given above seem to differ in terms of pinpointing the exact referent of the noun *problem* (something or someone, an issue or question, a matter or situation). What is similar about all of them, though, is that the semantic content of the noun implies the presence of the causative seme. So, it is possible to predict that the valence corresponding to the CAUSE relation should possess the highest probability rate (PR1 valence).

This prediction is strongly supported by the statistical analysis of the collected compound data. In 100 out of 117 compounds (86.2%) comprising the *N + problem* CWF the CAUSE relation connects the components, e.g. *health problem, waste problem, refugee problem, design problem*. The modifier component is understood as something or someone that causes problems, which means that although the relational information is targeted at the head, the semantic focus is on the modifier (which is noted to be typical of compounds in which the CAUSE relation is realised). Due to this kind of profiling of the meaning of the modifier element, i.e. the fact that something is wrong with it, it is implied even when it is expressed by a semantically neutral noun, e.g. *water, traffic, safety*. In this case the structure of the concept PROBLEM invokes our world knowledge about the modifier concept, as well as the negative associations that can be connected with this. For example, in the case with *employment problem*, it is not the employment opportunities that cause a problem but the lack of those. The same situation (when there is a lack/shortage of what

[107] The process appears to be quite the opposite; that is the adjectives of this group motivate the formation of nouns by means of conversion, e.g. *viral* (a piece of video circulated on the Internet), *social* (an informal social gathering), *facial* (beauty treatment for a face), *dental* (a dental consonant), etc.

the modifier refers to) can be seen in compounds *health problem, water problem, storage problem, mobility problem, money problem*, etc. The situation is quite the opposite for compounds like *traffic problem, crime problem, drink problem, drug problem, free rider problem, odour problem, waste problem*, etc., in which the abundance (or the existence) of what stands behind the modifier interrupts the normal course of life. Generally, the modifiers referring to objects and phenomena that are essential for normal functioning of life (*health, money, work, language*) are understood as not being enough, when combined with the head *problem*; and the ones that refer to negative aspects of life (*waste, anxiety, pollution, congestion*) are profiled by the head *problem* in the opposite way.[108] At the same time, it is necessary to point out that the implied abundance or lack of something, which is the cause of a problem, can be dependent on a number of extralinguistic factors. For example, *rain problem* in summer months will most likely refer to a lack of rain, whereas the same compound will probably refer to the opposite in winter months; *traffic problem* is conventionally understood as abundance of traffic that causes congestion of roads, which might be different if said by a person living in a remote area.

The above examples demonstrate that the semantics of the modifier as well as our pragmatic knowledge have considerable influence on our understanding of *N + problem* compounds. However, it is clearly seen that the conceptual structure of the head noun *problem* does not seem to suffer from alterations to the same degree as the conceptual structure of the modifier noun does. Perceiving the modifier noun, even when it is semantically neutral (e.g. *rain, water, traffic*), as a potential (or real) cause of problems arises from the strong negative associations with the concept PROBLEM. These associative links are directed at the modifier concept in such a way that we have to access our background world knowledge in order to calculate possible negative effects that the modifier can have. This suggestion implies that presuming that either the head or the modifier have a leading role in the way we operate with compounds may be an overly simplistic way of presenting the approach of the current thesis and that the process appears to be much more complicated than that. It was suggested above that the conceptual structure of the head has more weight in assigning the reading a compound receives, which can be clearly seen in *N + problem* CWF. Although a number of experimental studies provide evidence that the modifier concept is more influential in processing of compounds, it would be premature to assume that the same is true for creating them. It seems more reasonable to assume that different processes have

[108] It is not clear how to classify *tax problem* in this respect. The construal of the meaning probably depends on the context and the viewpoint, i.e. the government's tax problem should be different from the taxpayer's tax problem.

different directions and the modifier and the head have different roles depending on whether compounds are perceived or (re)produced.

The collected data of *N + problem* compounds also demonstrate the use of the LOCATION relation (*world problem, knee problem*) and the POSSESSION relation (*family problem*). The valences corresponding to these relations are not easily predictable from the semantics of the constituent and are not used very productively (8.6% and 5.2% respectively), and are therefore assigned the low probability rate (PR3). A closer look at the compounds comprising both of these groups still makes it possible to see the presence of the causative meaning profiled by the head *problem*. For example, if we have a knee problem, we understand that something is wrong with our knee and this 'something' results in the disruption of our normal functioning. In the case with *family problem*, we can say that 'a family has problems' or 'somebody has problems with their family', which also implies that the members of the family are affected by this. This can be explained by the strong associative ties in the head concept PROBLEM that imply the presence of the CAUSE/RESULT semes even in compounds with different relational meanings.

5.4 Compounds with a low to moderate preference for one position

5.4.1 Introductory comments

The group of compounds that demonstrate a low to moderate preference for being used as a modifier or a head in an N+N sequence comprises items with 51-79% preference for the relevant position. This group is the largest, with 54 CWFs (26 modifier CWFs and 28 head CWFs); however, in 38 out of 54 cases a constituent's preference for one position is ≥ 60%. In the analysis that follows we will look into CWFs of nouns whose count for the preference for one position is not so pronounced as for the previous two groups. For the analysis of the modifier CWFs I will look into the use of the elements *court* and *wind*, and for the analysis of the head CWFs I will discuss the CWFs of the elements *product* and *disease*. The discussion of each CWF will follow the outline used for the other groups.

5.4.1.1 *court + N* compounds: Miscellaneous notes

The collected data contains 38 examples of *court + N* compounds and 15 *N + court* compounds, which means *court* demonstrates an overall preference for the modifier position (72%). The meaning of *court* in the collected compound data refers to that of 'a court of law', i.e. 'a body of people presided over by a judge, judges or magistrates, and acting as a tribunal in civil and criminal cases', as well as 'the place where the court meets' (Oxford online 2012).

5.4.1.2 *court + N* compounds: Productivity of the semantic relation

The concept COURT can be described as a 'scenario concept'. In the case of concepts of this type we are dealing with a whole situation that has participants (judges, jury, defendants, attorneys, etc.), a prescribed place and procedure, possible outcomes, etc. Not much of this information is reflected in the dictionary definition, which is not surprising, because the concept is loaded with the information that is difficult to account for in the framework of a dictionary entry. Moreover, the semantic content of the word depends on the cultural, social, national, etc. factors; and even within the English-speaking world there will be variations, since the scenarios of the court procedure differ from one territorial administrative unit (country, state, county, province) to another.[109]

The suggestion that the concept COURT belongs to scenario concepts also implies that it is heavily dependent on knowledge of an extralinguistic nature, mainly on what we know about how things work in real life. This knowledge comes from our experiences, feelings, evaluations, associations, etc. Extralinguistic knowledge that is associated but not included in the meaning of the word is believed to serve as a base for the extension of the valence structure, making a larger number of valences available for activation. However, it is still assumed that the semantic content of a noun allows us to predict the valences of the highest probability of realisation based on the linguistic (semantic) information based on the dictionary definition.

Based on the dictionary definition, it seems that the active valences for the noun *court* should be the valence realised as the PURPOSE relation, which is expected to be realised in compounds describing the kind of affairs a court is dealing with (civil or criminal). It is also suggested that the realisation of the COMPOSITION relation is possible in compounds that specify the participants involved in the scenario (jury, magistrates, etc.).

The suggestions above do not seem to hold for the modifier CWF with the element *court*. Although the valence corresponding to the PURPOSE relation is activated, the relation is not very frequent (13.2% - PR2). The compounds in this relational group do not refer to *court* as an organisation that serves a certain purpose; quite the opposite happens: the PURPOSE relation is directed at the modifier element specifying it as the recipient of the relational information. At the same time, looking at the *N + court* CWF (*youth court, labour court, divorce court*, etc.), we can see

[109] Scenario concepts can be opposed to the so-called 'core concepts', or concepts possessing 'absolute cue reliability' (Ryder 1994: 81-82). Core words consistently project the same meaning irrespective of what they are paired with, e.g. the word *box* realises the meaning 'container'.

that the relational link realised here goes in line with the prediction above. This mismatch supports our hypothesis that it should be possible to predict the relational reading of a compound based on the semantic content of the head noun and not the modifier.

Based on the statistics, the most frequent relation in the *court + N* CWF is the LOCATION 2 relation (34.2% - PR1), e.g. *court office, court trial, court battle*. The LOCATION 2 relation is directed at the modifier, and in this case the element *court* is profiled as space where something takes place or is found. The necessity to point out the directionality here is dictated by the fact that the LOCATION relational group also contains compounds in which the directionality of the relation is different, e.g. *courtroom*. In this subgroup the head noun is profiled as a place, whereas the modifier noun is understood as an event or a process. The change in the meaning of the modifier noun is reflected in (or maybe caused by) the directionality of the semantic link. However, following the opinion accepted in this research, it is believed that such changes should be attributed to the abstract schema of the head noun, which predetermines the way in which the modifier fits into the construction.

The other most frequent relations in the *court + N* CWF are the SOURCE relation (18.4% - PR2), e.g. *court decision, court order, court summons* and POSSESSION (13.2% - PR2), e.g. *court jurisdiction, court school*. The compounds of the first relational group reflect the result of the process denoted by the modifier, whereas the compounds of the second group denote attributes that the modifier concept possesses.

PR3 valences are realised as the TOPIC (*Court Act*), INSTRUMENT (*court system*) and TIME (*court time*) relations. The realisation of these valences does not appear to be a productive process but the fact that a variety of valences are realised in the *court + N* CWF is important, as it supports the idea about the possibility of any valence to be realised as a semantic relation in a noun concept.

5.4.1.3 *N + product* compounds: Miscellaneous notes

The collected CWFs with the element *product* include 40 units in which *product* is used as a modifier, and 71 units utilising the noun as a head. The preference for the head position is classified as relatively low (about 64%); however, it is clear that *N + product* compounds are used more productively than *product + N* compounds, with almost twice the number of units with *court* used as the head (at least in the collected corpus). The size of the modifier CWF also allows for some comparisons in the patterns.

The noun *product* is more or less unambiguously defined across dictionaries (Longman online 2012, Oxford online 2012) as an article or substance that is grown or manufactured and is often intended for sale. Webster and Collins dictionaries provide a broader definition: 'something produced', which allows for inclusion of more abstract items like a service, the result of one's creative work (a film, a theatre performance, a book), the result of some experience, situation or action, etc.

The compounds in which the other meaning of the word *product* is realised, i.e. 'the number you get by multiplying two or more numbers in mathematics', were not included in the corpus because of the specificity of their meaning.

5.4.1.4 *N + product* compounds: Productivity of the semantic relation

The semantic content of the noun *product* is suggestive of the strong presence of a number of active valences that the word can realise in N+N combinations. The definition implies that we should expect the existence of the subject or a producer, what the produced item is made from and the recipient or the expected result of the act of producing. This, in its turn suggests the activity of valences that can be realised as the PURPOSE relation, the COMPOSITION relation and the SOURCE relation. Although the broad definition 'something produced' given in the Collins and Webster dictionaries makes the above predictions look rather speculative, more precise definitions (as given in Oxford and Longman dictionaries) allow for a suggestion that the discussed valences are part of the word's meaning. At the same time, it is also clear that a dictionary definition cannot be considered an ultimate source of information about a word's semantic content and its aim is only to capture the most salient semantic features that allow for describing the meaning more or less accurately.

Whatever definition of the noun *product* we choose, there is a connection between the verb *to produce* and the noun *product*. The latter probably evolved as a nominalisation of the former, since the form *produce* was borrowed from Latin *producere* (bring forth), which is a verb. Therefore, this suggests a connection between the arguments of the verb and the valences of its nominalisation.

The analysis of the data shows that the above prediction can be considered rather accurate, with the PURPOSE relation being a realisation of a PR1 valence (*household product, consumer product, beauty product*). The modifier nouns in the PURPOSE relational group imply the reference to the intended recipient of some product or the location where a product can be used (12 out of 29

compounds), e.g. *consumer product, hair product, office products, home product, garden product*; or an activity that a product can be used for (13 out of 29 compounds), e.g. *cleaning product, building product, styling product, communications product, storage product, skincare product*, etc. In the latter case the PURPOSE relation also bears a shade of instrumentality, when the head is understood as 'something (unspecified) used for performing the activity'. The reference to an intended result is also possible, e.g. *beauty product, safety product*, but it does not appear often in the data (four cases out of 29). Although the range of meanings that are felt to be present in this relational group is rather wide, they are still considered to be manifestations of the same relational meaning. The differences can be explained by the broadness of the PURPOSE category.

The second most common relation realised in the *N + product* CWF is the COMPOSITION relation, with 26.8% of occurrences, which means that the corresponding valence can be assigned PR2 status. The first component in this group is generally understood as the material/substance from which a product is made, e.g. *milk product, tobacco product, fish product, blood product*. In one of the examples in this relational group the modifier does not provide the direct reference to the material. In the case of *animal products*, the first component is used conventionally to describe any material taken from the body of animals, including milk, fat, eggs, meat as well as products like rennet and meat or fish by-products, and other products received from animals like beta carotene, glycerine, lipase, etc. In this case the referent of the modifier is rather generic but the compositional meaning is felt to be preserved.

The other PR2 valence is realised as the SOURCE relation (11.3%), e.g. *farm product, forest product*, etc. In this group the meaning conveyed by the modifier is that of the subject/producer of the product and conveys the meaning 'originates from'.

All of the three most frequent relations were predicted from the semantic content of the noun *product* used in the head position. If we compare the relational information of the *N + product* CWF with the relational information of the *product + N* CWF, we can see that these readings are not so frequent and the valences that are active for *product* as a modifier are not easily predictable from the semantic content of the noun. For example, based on the analysis, PR1 valence for *product* in the modifier position is realised as the POSSESSION relation, where the modifier is viewed as a possessor of something denoted by the head, e.g. *product quality, product design, product title*, etc. This information is not immediately included in the meaning of the word, but is connected with our knowledge of extralinguistic reality. One objection could be based on the fact that the PURPOSE relation is rather actively used in the *product + N* CWF (PR2 valence –

20%). However, the PURPOSE relation is directed at the modifier (*product* ← *N2*), thus making it the target, unlike the combinations in the *N + product* PURPOSE relational group, where the head is targeted at something denoted by the modifier (*N1* ← *product*); cf. *product division, product programme* vs. *health product, garden product*. So we can see that although the semantic link is the same, its direction in relation to the noun *product* in different CWFs is different and the information about a product being an aim or a target of something is not part of the meaning of the word but part of what we know about it.

In terms of the directionality of the relation, the situation is a bit different with the POSSESSION groups in both CWFs; cf. *product quality – quality product, product price – value product*. In both *product + N* and *N + product* CWFs the POSSESSION relation suggests that the constituent *product* is viewed as a possessor: *product quality* (product HAS N2), *quality product* (product HAS N1). Here we seem to be dealing with the situation when the relational information does not seem to provide the distinction in the reading of compounds but the difference in meaning is determined by grammar and the positioning of the elements, which may indicate that in one case we are dealing with *quality* and in the other one with *product*. However, the POSSESSION relation is not very frequent for *N + product* compounds (2.8% - PR3 valence), which suggests that understanding the concept PRODUCT as a possessor of something is not one of its most salient features. At the same time the fact that this relation is realised, as well as other relations (LOCATION – *niche product*, INSTRUMENT – *technology products*, ESSIVE – *hardware product*, CAUSE – *breakdown product*, TIME – *end product*) supports the idea about the potential presence of any valence in the meaning of a given noun concept.

5.4.1.5 *wind + N* compounds: Miscellaneous notes

The preference for the modifier position demonstrated by the key element *wind* can be characterised as medium with 75.8% of cases. The collected data suggests that the use of the key element in the modifier position is productive with the majority of *wind + N* compounds being low-frequency items (at least in the BNC). The collected *wind* as head CWF is not very large, with 15 items comprising it.

The two main meanings of the word *wind* are described as a 'perceptible natural movement of the air, especially in the form of a current of air blowing from a particular direction' and 'breath as needed in physical exertion, speech, playing an instrument, etc.' (Oxford online 2012). Interestingly, Oxford and Collins dictionaries (Oxford online 2012, Collins online 2012) list these meanings as different, Webster (Webster online 2012) lists the two as constituents of one general

meaning, and Longman (Longman online 2012) does not have such a division at all, providing the most general definition 'moving air'. The key element *wind* in the collected data realises both of the meanings. However, compounds in which the word *wind* realises more specific meanings (which are believed to be metaphor-based), i.e. wind in the stomach, a trend or force, etc. are not included in the analysis. At the same time, compounds in which the element wind realises the meaning relating to wind instruments, e.g. *wind band, wind music*, are included in the corpus, since it is believed that the main components of meaning of the word are preserved (though it is understood that the meaning is extended).

5.4.1.6 *wind + N* compounds: Productivity of the semantic relation

The semantic content of the noun *wind* (as outlined in dictionaries) is strongly connected with the idea of movement, which suggests the possibility of predication that is typical in descriptions of actions of movement. So, based on the word's meaning, the possible active valences should be those that imply LOCATION[110] and TIME. The activation of these valences is strong in the *N + wind* CWF, both of which can be assigned PR1 status for the noun used in the head position, but not in the modifier position.

As the analysis shows, no relation in the *wind + N* CWF can be described as a realisation of PR1 valence. Three of the most frequent relations (realisations of PR2 valences) in this CWF are POSSESSION (*wind speed, wind direction*), INSTRUMENT (*wind farm, wind music*) and SOURCE (*wind energy, wind advantage*). As has been suggested above during the discussion of other modifier CWFs, the relational information for the modifier does not seem to be easily predictable from the semantic content, but is heavily dependent on background knowledge. This suggestion holds for the *wind + N* CWF too, and the realisation of the relations mentioned above is not directly connected to the semantic content of the modifier concept.

Looking more closely at the relational groups in which PR2 valences are activated we can notice the following. The head nouns in the POSSESSION relational group describe the characteristics of the natural phenomenon denoted by the modifier. Here *wind* is profiled as a possessor of certain characteristic features, like speed, strength, velocity, angle, etc. Although we know that wind as a phenomenon has such characteristics, this knowledge is not explicitly present in the semantic content of the word, but is stored in the background. This kind of knowledge is felt to be of

[110] The list of Levi's semantic primitives does not include the DIRECTION relation or SOURCE OF DIRECTION, these are included in the more general LOCATION category.

encyclopaedic nature and to be somewhat different from the knowledge whose involvement is necessary for producing and perceiving most of the compounds in which the INSTRUMENT relation construes the sense, like *wind turbine* or *windpipe*. In cases like these we deal with naïve, superficial perception of the objects described by compounds, i.e. objects in which wind is used as an instrument.

The INSTRUMENT relational group is worth discussing in a bit more detail, since it can be considered that the modifier in this relational group realises different meanings. That is, *wind* in *wind farm* is understood as the natural phenomenon, as opposed to *wind* in *wind band*, where *wind band* is aphetic for *wind instrument band*. However, looking at examples of this type from a different perspective, it can be seen that for musical instruments in compounds like *wind band* or *wind orchestra* the most salient characteristics are similar to the meaning of the noun *wind* in the reference to the natural phenomenon. Although the source of the wind is different, the general outline of the characteristics remains the same. The meaning of the modifier in examples like *wind music* or *wind band* can be considered a metonymical extension of the general meaning of the word, in which the natural movement of air is compared to the movement of air that is produced for certain purposes (in the discussed case, for powering some musical instruments). With years of conventional usage, the metonymy has lost its novelty and become stale. At the same time the transparency of the metonymical extension appears to be preserved and the instrumental meaning used as the basis for the metaphor is easily accessible. The accessibility/transparency of instrumental meaning in compounds like *wind music, wind band, wind orchestra*, etc. may attest to the existence and regularity of compounds like *wind instrument, windpipe, windbag* (in its literal sense), which use the metonymical extension 'first-hand' and can be considered prototypes that encourage the formation of the '*wind* for *wind instruments*' paradigm. 'Second-hand' extensions in which *wind* is understood as a kind of a conceptual truncation of the compound *wind instrument*, as in *wind band*, are believed to be cases of partial metonymy, in which the modifier noun serves as means of accessing the full referent.

The other PR2 valence (21.7%) is realised as the SOURCE relation, e.g. *wind conditions, wind energy, wind noise*. The SOURCE relational group is opposed to the group in which the modifier concept is profiled as the CAUSE of something denoted by the head noun, e.g. *wind damage, wind erosion*. These relational groups are felt to be quite similar; however, the head nouns in the causative group often denote negative effects of the wind (based on the collected corpus). However, it would be wrong to claim that the presence of a negative shade of meaning should be

taken as a landmark for distinguishing between the CAUSE and SOURCE relations, since there are still cases for which this does not hold true, e.g. *wind pollination*. At the same time, it is clear that these two relational concepts are very close in terms of the meaning they project.

The last two valences (PR3) are realised as PURPOSE (*wind window*) and NEG PURPOSE (*wind straps*) relations. The distinction between these two relations is based on the FOR vs. AGAINST opposition, with the element *wind* in the compounds of the second relational group being conceptualised as something negative (which might be context-dependent).

> All of these generalisations depend of course on a constant wind speed sufficient to fly the kite within its designed **wind window**.

We have seen that the modifier in the *wind + N* CWF demonstrates activation of a number of valences, the majority of which are not easily deducible from the semantics of the modifier concept. The *wind + N* CWF does not seem to demonstrate a strong preference for realising one semantic relation, but at the same time some relational links are felt to be similar, e.g. SOURCE and CAUSE.

5.4.1.7 *N + disease* compounds: Miscellaneous notes

Based on the collected data, the noun *disease* demonstrates a medium preference for being used in the head position, with 58 out of 80 compounds comprising the *N + disease* CWF. The majority of compounds (39 out of 58) in the head CWF have a relatively low frequency of occurrence in the BNC (≤20), which may serve as an indicator of the productive use of the element in the head position. However, we cannot say that the use of the *disease + N* pattern is unproductive. In fact, a larger proportion of *disease + N* than *N + disease* compounds have a low frequency of occurrence in the BNC, which means that the productivity of the key element as a head or as a modifier is somewhat questionable. Since this research takes on the synchronic approach, making predictions about how the situation may change in the future is not one of the questions that I have attempted to answer here; and we will concentrate on the analysis of the noun *disease* as the head CWF, since it appears to be utilised as the head of N+N sequences more actively now.

The meaning of the noun *disease* as described by the Oxford dictionary (Oxford online 2012) is that of 'a disorder of structure or function in a human, animal, or plant, especially one that produces specific symptoms or that affects a specific location and is not simply a direct result of physical injury'. The definitions from other dictionaries add some other specific aspects of meaning that allow for representing the semantic content with more precision; e.g. 'caused by

bacteria or infection' (Collins online 2012), 'impairs normal functioning' (Webster online 2012). As will be seen below, the analysis of the collected compound data shows that these elements of meaning are well represented in the *N + disease* CWF.

5.4.1.8 *N + disease* compounds: Productivity of the semantic relation

Based on the definitions above, the semantic content of the noun *disease* implies the activity of the valences that outline the location of a disease, its bearer and cause. This means that we can predict that the LOCATION, POSSESSION and CAUSE relations should be the most frequent in N+N combinations with the key element *disease*. The statistical analysis of the semantics of the compounds supports these predictions for the *N + disease* CWF rather accurately. The LOCATION relation is assigned PR1 status, with 50% of compounds in this CWF realising this relation, e.g. *liver disease, eye disease, chest disease*. The modifier noun in this relational group denotes a part of the body that is characterised as space or location of some sort in which a disease is found. It can be argued that this group may be categorised as the POSSESSION relation, with the interpretation 'N1 HAS *disease*'; and in this case the modifier noun is conceptualised as a possessor of a disease. However, the compounds of this relational group are felt to be different from the ones that were analysed as the POSSESSION group, e.g. *family disease, potato disease, fish disease*. The modifier in the POSSESSION group denotes items that are viewed as a whole (*family, tomato, cattle*, etc.), whereas the modifiers in the LOCATION relational group refer to items that are parts of animals or plants (*eye, ear, gum, foot*, etc.). The only compound in the LOCATION relational group whose modifier is somewhat difficult to describe as being a part of the body is *blood disease*. However, based on the dictionary definition, it is possible to claim that *blood* here may still be perceived as some sort of space, in which malfunctioning caused by viruses or bacteria takes place.

The POSSESSION relational group accounts for 22.4% of occurrences and the CAUSE relational group makes up 20.7% of cases. The valences realised as the POSSESSION and CAUSE relation are assigned PR2 status, and the activity of these valences can be predicted from the semantic content of the noun *disease*. As mentioned above, the modifier in the POSESSION group denotes a whole unit that is characterised as a possessor of a disease. The connection of the elements by means of the POSSESSION relation is also noted in the CWF where the element *disease* is used as a modifier, e.g. *disease rate, disease conditions, disease extent*. It can be noticed, though, that due to the directionality of the relation, the element *disease* is conceptualised as a possessor of some qualities denoted by the head noun (*disease* HAS N2), which makes this group different from the POSSESSION group of *N + disease* compounds, where N1 is understood as a possessor (*tomato*

HAS *disease, elm* HAS *disease, plant* HAS *disease*). Comparing the two POSSESSION groups, we can see that understanding the concept DISEASE as a characteristic of something/someone is more typical and conventional than understanding DISEASE as being characterised. That is, it is easier for us to perceive something as having a disease, than to perceive the disease as possessing certain characteristics. This somewhat naïve perception is reflected in the semantic content of the concept, which makes it possible to predict the former and not the latter.

The CAUSE relational group contains compounds whose modifier names the reason for the disease, e.g. *fungus disease, virus disease, (X) deficiency disease*. The modifier in this group may also be expressed by an *-ing* nominalisation that outlines an activity causing a disease, e.g. *smoking disease, slimming disease, diving disease*. Quite often the cause of the disease named by the modifier noun can also be viewed as a result, e.g. *stone disease, ulcer disease, gas bubble disease*. For example, in the case with *stone disease*, we are talking about a medical condition that is characterised by the formation of calculi (stones) in the urinary tract and, in the naïve understanding of the matter; the calculi are the reason for the medical condition.

The causative group is also present in the *disease + N* CWF, e.g. *disease cells, disease cultures, disease organisms*. The directionality of the relation in the modifier family (N2 CAUSE *disease*) suggests profiling the key element *disease* in the same way as in the head family (N2 CAUSE *disease*); however the grammatical order of the elements is different. As was the case with the POSSESSION relational groups in *product + N* and *N + product* CWFs, here we are dealing with the differences in meaning as determined by grammar. It is worth noting that the number of compounds realising the CAUSE relation in the modifier CWF is not so large (four compounds out of 22 comprising the modifier CWF), which suggests that this relation does not have a high PR for *disease* as a modifier, despite being highly predictable from the semantic content of the noun.

The other valences that are assigned PR3 status are realised as the ESSIVE (*killer disease*) and the TIME (*childhood disease*) relations. These valences are not actively used and are not easily predictable from the semantic content of the noun *disease*.

5.5 Summary of Chapter 5

In the course of the analysis of the most representative CWFs comprising the collected corpus I have looked at the constituents that demonstrate different degrees of preference for being used as the head or the modifier of an N+N sequence. In the course of the discussion I have attempted to show that the semantic relations realised in English N+N compounds can be predicted from what we know about the constituents as lexical items and as labels naming the objects and phenomena of extralinguistic reality. The possibility of predicting the semantic relations is strongly connected with the role the constituent plays in a sequence. The semantic links in compounds are not easily predictable when the analysis concentrates on the modifier element and the analysis of the semantics of the modifier fails to provide relevant information. I have shown that predictability of the relational information realised by the noun concept when the modifier position is taken as a standpoint, in many respects relies on the knowledge of extralinguistic nature. This includes extralinguistic knowledge like our world knowledge, personal experience, emotional values, etc. as well as the inner statistical knowledge about how the modifier is used in constructions of this type and which serves as the basis for analogy. The semantic content of the modifier concept does not appear to play the leading role in defining which semantic relation can be realised by the modifier constituent.

The situation is different if the analysis approaches the semantics of compounds from the position of the head. The semantic content of the head constituent (based on a dictionary entry) is believed to be a rather strong predictor of the relational information that is projected by the whole compound.

I have attempted to explain how the valence structure of a noun concept allows for activation of different valences and their realisation as different semantic relations. The activity of certain valences is determined by a number of factors, of which the main ones are the semantics of the noun concept and its position in a compound. The valences that have a higher PR when a noun is used as the head of the construction are usually connected with the semantic content of the noun, and the valences with the lower predictability of realisation are connected with other knowledge. The valences with the lower PR are connected with the information that is peripheral to the meaning of the concept. The analysis of valences that appear to be most active when a noun is used in the modifier position suggests that their link with the core semantic content is not essential and any valence can be realised as a predominant semantic relation. A lesser degree of dependency of the modifier on the semantic content of the noun in terms of activating one of the

more predictable valences can be supported by the fact that the highly predictable relation does not always surface as active and has a marginal number of cases. For example, in the case with *court + N* CWF the formation of compounds (and the utilisation of relational information) is more reliant on analogy and not on the semantic content of the noun, at least from a synchronic perspective. Another example is *family + N* CWF, where the most frequent relation is that of POSSESSION. However, the POSSESSION relation does not appear to be in direct correlation with any of the meanings of the word. The consistent realisation of this relation despite the lack of the direct connection with the semantics of the noun suggests that understanding of *family* as 'a possessor' is tied to what we know about families rather than to the meaning of the word.

These findings challenge the claim about the leading role of the modifier in defining the relational reading a compound receives and the semantic weight of the modifier in an N+N sequence. Although there is no denying that the modifier concept changes the meaning of the head concept dramatically, and this can even be reflected in the semantic relations (*health issue* vs. *family issue*), it was shown that the change of meaning of the semantic relation is connected with the semantics of the head. For example, we could see that the POSSESSION relation is not a typical one based on the semantics of the modifier *family;* however, it has the highest probability rate based on the analysis of the collected data.

Another argument that supports this is connected with the fact that the semantics of the modifier is altered and accommodated to suit the abstract schema of the head noun, whereas the head element generally preserves its general meaning in endocentric N+N compounds. For example, in compounds *health centre* and *community centre* the head refers to a place, even though the mental image of the place differs in each case. In *health service* and *health food* the referents of the modifiers are obscure and the meanings projected by them are inherently different. It is assumed that the possibility of projecting different meanings by a modifier element (within the boundaries of possible meanings a noun concept can realise) is related to the partial loss of the semantics by the modifier noun, which is related to the function of delimiting/specifying the meaning of the head that the modifier fulfils. This suggestion is supported by the experiments conducted by Libben, Gibson, Yoon and Sandra (2003), whose empirical data demonstrate that the semantics of the head is more important for processing compounds.

In the course of the analysis of constituents whose preference for the head or modifier position is less strong, we compared the semantics of the relations within modifier and head families for one and the same constituent. The analysis provided more evidence that predicting the realisation of

the semantic relations based on the semantic content of a noun is more accurate for the noun used as a head rather than as a modifier.

In the course of the analysis of individual constituent families I have also attempted to show that compounds with larger family sizes tend to instantiate fewer semantic relations than compounds with smaller constituent family sizes. This means that there is a connection between the productivity of a noun concept as a constituent of a compound and the concentration on one semantic relation by this concept in the constituent family.

It has also been shown that in general modifier CWFs have a lower consistency in terms of realising one relation as compared to the head CWFs. Modifier CWFs seem to be consistent only when the semantics of the noun suggests that it can be viewed as a core concept, as is the case for the constituent *future*, whose modifier CWF demonstrates the strong concentration on the temporal relation.

6. Conclusion

6.1 Introductory comments

The aim of this chapter is to provide a general overview of the results obtained in the corpus study of English N+N compounded structures. The analysis of the collected data has revealed several significant trends in the formation of compounds in English. The trends noted in the course of this investigation demonstrate how morphological, semantic and conceptual information is connected in the process of forming conceptual combinations which are structurally expressed by N+N sequences in English.

6.1.1 Summary of the findings

This thesis addresses the questions of formation of English N+N sequences. In my research I have looked at the issues of morphological and semantic productivity and attempted to provide explanations for the trends that were noted in the course of the analysis. This section outlines the main findings of this study and their implications.

1. *The current research demonstrated that there is a preference for a noun to be used more productively in only one position (modifier or head).*

One of the hypotheses that the analysis checked is whether, as a constituent of a compound, a lexeme is productive in one position and not the other. The collection of the corpus of 7332 compounds comprising 97 constituent families allowed for a statistical check on the productivity patterns in the use of constituents of a compound. The analysis of the data clearly indicates a negative correlation between the sizes of the modifier and head families for one constituent, i.e. with the increase of the head family size for a given noun, the size of the modifier family for the same noun decreases. These results suggest that any given noun is used more productively either as a modifier or as the head of the compounded structure. This result supports Baayen's (2010) claim that that in the majority of cases the constituents of lexicalised compounds are position-bound. The absolute majority of compounds used in the current study are not lexicalised, and since this holds true for nonlexicalised compounds too, the results obtained suggest that in the process of forming a new compound on the level of structural representation we follow a certain pattern that predetermines the use of a noun as a head or modifier. The factors that influence the productive use of the noun concept in the head or modifier positions are not always clear. Based on the data, the opinion that a limited number of lexicalised compounds within a CWF can

promote further formation of structurally similar items by means of analogy (Booij 2007) appears to be somewhat premature.[111]

It is also worth pointing out that the degree of preference for being used in one position differs for different nouns and ranges from absolute (100%) to low (≥51%), which allows for the suggestion that morphological productivity of an element in one position is gradient. However, the size of the collected corpus does not allow for making unambiguous predictions concerning the features of noun concepts that may be responsible for this. One of the preliminary speculations that can be put forward concerns the morphological properties of the noun. For instance, items that are adjective-like, e.g. *future*, *animal*, *chocolate*, *lemon*, have a higher probability of occurring in the roles typical of an adjective (attributive role in our case) and, therefore, are more often used in the modifier position than in the head position. The analysis of a larger data set might help shed light on this issue.

Another possible explanation of why some noun concepts are head-oriented (e.g. *bill, matter, period, order*, etc.) and others are modifier-oriented (e.g. *disease, police, property, trade*) can be formulated as follows. Based on the analysis of the collected data, analogy in word-formation should still be the most influential factor for an element's positional preference. However, the results imply that the extension of one paradigm (modifier or head) but not the other is difficult to prove if we accept Booij's (2007) suggestion that only one item should serve as a prototype that motivates further formations. Although the latter possibility should not be excluded, the collected data demonstrates that pinpointing one case that may have served as a prototype is only possible for a limited number of CWFs (mostly novel ones), and such criteria as lexicalisation, frequency of occurrence and time of coming into regular use do not seem to be significant. For example, if we look at the constituent families for the element *station*, the existence of the lexicalised item *stationmaster*, which has a relatively long history of existence in the language (Webster online 2012), does not seem to promote the productive use of the element in the modifier position. Quite the opposite, the family size of the *N + station* suggests that the element should be considered head-oriented with a very high preference for the head position. Another example is *girlfriend*, which has a very high frequency count in the BNC (1475). However, the use of the element *girl* in the modifier position can hardly be considered very productive with only 11 compounds comprising the modifier CWF, as opposed to 58 compounds in which *girl* is used as the head. Based on the analysis of the collected corpus, it is assumed that the productivity of a given

[111] See Section 2.6.1 for discussion.

element in a certain position is driven by analogy on the paradigmatic level, i.e. the size of the constituent family is a more important factor for the extension of one paradigm over the other.

This finding implies that we can talk about the productivity of a noun when used only as a head or modifying constituent in a compound, which means that putting two nouns together in order to denote objects and phenomena of real life should be considered a morphological process rather than a syntactic one.

2. *Productive use of a constituent of a compound is also connected with the productivity of the semantic relation that is realised by this constituent in compounded structures.*

Ryder (1994) observes that conceptual combinations with some nouns are consistent in realising one relational meaning. For example, nouns like *box* are normally used in their primary meaning 'container' and the semantic relation in N+N compounds with this noun appears to be the same for combinations like *shoe box*, *chocolate box*, *tool box*, etc. Ryder (1994) calls concepts like that 'core concepts' and also points out that not all noun concepts demonstrate an equal degree of 'coreness'.

This research has looked into the semantic relations realised by noun concepts as constituents of compounds in order to check whether there is an overall preference for a limited number of semantic relations to be realised in a constituent family. The statistical analysis of the collected data described in Chapter 3 demonstrates that this is true. It has also been found that the size of the constituent family can be considered a strong predictor of the degree to which a constituent concentrates on one semantic relation. The results obtained demonstrate that the possibility of instantiation of one particular relation over the others increases as the family size increases, notwithstanding a logical supposition that large CWFs should have a wider distribution of semantic relations. This tendency is slightly stronger for head CWFs than for modifier CWFs. The consistency in realising one relation varies for different concepts; but, as suggested by the statistical analysis, the preference for one relation is not random and is consistent in the collected corpus. This means that there should be a connection between the productivity of a noun as a constituent of a compound and the consistent realisation of one semantic relation by this noun within a constituent family.

The analysis of the results allowed for the suggestion that the complexity of the semantics of noun concepts can be considered an influential factor for a word to demonstrate a) a less strong preference for one position and b) a less strong preference for realising one relation consistently. For example, concepts like *conference, business* and *community* are opposed to concepts like *tree,*

lemon and *station* in terms of complexity of their semantic content, which is also revealed in the number of semantic relations realised in CWFs containing these concepts. One of the criteria for accounting for the complexity of the concept is the word's abstract or concrete referent, with nouns naming abstract phenomena being more complex than nouns naming concrete objects.

It is necessary to point out that the meaning of a word, as listed in dictionaries, is rather the tip of an iceberg of what a word stands for and does not include all the emotional, experiential and pragmatic components that a word possesses. In this research the term 'semantic content' is used to talk about the word's meaning as listed in dictionaries.

The analysis of the words' semantic content was used to check the possibility of predicting the semantic relations that a word is likely to realise as an element of a compound. It was noticed that the realisation of the semantic link is directly connected to the core semantics of the head noun in an N+N compound. The semantics of the head noun serves as a schema that allows for a modification of a certain type, which is what the modifier noun does. It has been shown that the most frequent relation realised in the analysed head CWFs can generally be predicted from the headword's meaning. For example, in the case of *N + problem* compounds, the meaning of the head, i.e. 'a situation, matter, or person that presents perplexity or difficulty', suggests that the causative seme is embedded in the semantic content of the noun. Not surprisingly, the CAUSE relation is the most frequent relation realised in this constituent family.

The core semantic content of the modifier does not seem to be utilised for the construal of meaning of a compound in the same way as the semantic content of the head constituent. This makes predicting the possible semantic relation in an N+N compound from the modifier's semantics difficult. As was pointed out in the course of the data analysis, the semantic relation is connected with the modifier via a different kind of knowledge, including experiential, encyclopaedic, pragmatic and statistical knowledge.

This finding has important implications for further research in this area. One of the suggestions discussed in this thesis concerns the use of analogy in the process of coining a compound. The results obtained in the course of the current study imply that the use of analogy in the formation of N+N compounds takes place on at least two levels of representation: structural and semantic. It also appears that there should be interaction between the levels of structural and semantic representation in the process of forming compounds. Further research in this direction would to be beneficial for a deeper understanding of the processes involved in the formation of compounds.

3. *The head and modifier constituents of an N+N sequence have different semantic weight in an N+N compound.*

One of the points that were discussed in the course of the analysis of the English data concerns the suggestion that relations connecting noun concepts should be viewed as realisations of valences that every noun possesses. As is argued here, the set of valences is the same for every noun concept; however, the probability of activation of different valences differs depending on the word's semantics. For example, the semantics of the noun *law* in the meaning 'a system of rules' suggests the activity of the valence that can be realised as the TOPIC relation (law ABOUT something). At the same time, based on the word's meaning, the valence that can be realised as the ESSIVE relation (as in *draft law*) is not very active. Therefore, the valence realised as the TOPIC relation should be more active than the valence corresponding to the ESSIVE relation. However, the latter is potentially present and can be used should the need arise. In the case with the noun *area* it is more likely to be conceptualised as a kind of space (as in *garden area*) than as a kind of cause (as in *risk area*). Thus the degree of probability of occurrence for the valences LOCATION and CAUSE is different in the case of the noun *area*. The degree of probability of different valences depends on a number of factors including the semantics of the concept, their position in the conceptual combination (and thus the role they play), frequency of occurrence, conventionality of usage, etc.

I suggested that the constituent whose semantics is more important for the meaning of the compound should demonstrate consistency in realising the same relation in the constituent family, and this relation should be predictable from the semantic content of the constituent concept. I looked at individual constituent families to see whether the most frequent relation or relations are tied to the semantic content of the constituent. In the course of the analysis it was noticed that the degree of potential activity of valences is reflected in the semantic relations realised in CWFs for elements used as heads. Chapter 4 discusses only a limited number of most representative cases; however, the analysis of the whole corpus demonstrates the same trend. Based on this, it is possible to predict the possibility of realisation of a certain semantic relation from the semantic content of a noun when it is used in the head position. This suggests a strong connection of the semantic relation with the semantic content of the head, which, in the absolute majority of cases, determines the overall reading an N+N sequence receives.

The situation is different for the active valences for nouns when used as modifiers. The analysis of relations that appear to be most frequently realised when a noun is used in the modifier position

suggests that their link with the core semantic content is not essential and any semantic relation can be realised as predominant (including those that are connected with the semantics of the modifier concept). A lesser degree of dependency of the modifier on the semantic content of the noun can be supported by the fact that the most highly predictable relation does not always surface as active and has a marginal number of cases. The comparison of the semantic relations realised in the head and modifier CWFs for a given element suggests that the valences that should be potentially active for a noun are not consistently realised as semantic relations when a word is used as a modifier in an N+N sequence. For example, the semantic content of the noun *property* suggests that the valence realised as the POSSESSION relation (N1 HAS *property*) should be predominant, which it is for the *N + property* CWF, but not for the *property + N* CWF. The relational information in the modifier CWFs appears to be random at first sight. On further consideration, however, one idea seems plausible.

The valences realised in modifier CWFs are not predictable from the semantic content of the noun, but are dependent on what we know about the concept this noun denotes. In the course of analysis of modifier CWFs, we have seen that the connection between the modifier concept and the realised relation is connected with the semantic content of the modifier noun indirectly. For example, in the case with the *community + N* CWF, the semantics of the key element suggests that the most active valences should be those realised as the COMPOSITION relation and the LOCATION relation. However, the former is not represented in the CWF with *community* as a modifier, and the latter has a low number of instantiations in the family. The predominant relations in the *community + N* CWF are those of PURPOSE (*community programme*), POSSESSION (*community power*), and TOPIC (*community research*). These relations are connected with the information that is peripheral to the semantic content and does not seem to be salient and/or obligatory. Compounds in which these relations are realised reflect our world knowledge about the concept COMMUNITY but this information is not directly connected with the semantic content of the concept.

These findings challenge the claim about the leading role of the modifier in defining the relational reading a compound receives and the semantic weight of the modifier in an N+N sequence. Although there is no denying that the modifier concept changes the meaning of the head concept dramatically, the change of meaning of the semantic relation seems to be connected with the semantics of the head.

I have tried to show that the head concept of N+N constructions seems to have a leading role in terms of determining the relational reading of the whole coinage. The head noun serves to provide the domain, whose modification is determined by the semantic schema of the noun concept that labels this domain. The choice of the modifier is dependent on the distinguishing features of this domain and the relational link is determined by the semantic content of the head concept.

The claim about the priority of the semantic content of the head noun for a compound to acquire a certain reading is based on the approach that views semantic relations as realisations of potentially existing valences, which are the same for every noun concept. It is suggested that the activity of valences for a given noun concept is determined by its semantic content, and that the most active valences have the highest probability of being realised as semantic relations in conceptual combinations. The analysis of the collected corpus of English compounds demonstrates a strong connection between the semantics of the head concept and the semantic relation realised in the head constituent family, which is not the case for modifier families. The influence that the modifier concept has on the relational reading of a given N+N sequence in English is not connected with the core semantic content of the modifier but to a larger degree depends on our extralinguistic experience with the concept denoted by the modifier. Our inner statistical knowledge about the ways in which the modifier noun can be used in the language is also an important factor that contributes to the ways of using a noun in the modifier position.

Although the modifier constituent does not seem to have much influence on the relational reading of a compound, its role in meaning construal of a compounded structure should not be underestimated. The purpose of the modifier is to specify the semantic content of the head in the way that is necessary for conveying the idea intended for communication. In the course of the analysis of the English compound data we could see that the meaning of the head noun may undergo considerable changes under the influence of the semantics of the modifier. For example, our understanding of *island* in *traffic island* is different from *desert island* in English.

6.1.2 Where do these findings take us?

In the course of the research a number of things that would be interesting for further investigation were detected. For example, some modifiers (e.g. *animal, future, family*) seem to behave more like adjectives than others and there are no clear-cut criteria that can be used to refer such items to any of the categories. In this thesis no distinction was made between adjective-like nouns and clear-cut nouns, however a comparative/contrastive study that would involve obtaining

psycholinguistic evidence may be beneficial for answering the question whether adjective-like nouns are different from clear nouns (cf. Bell 2011).

In the course of the data collection it was noticed that we may come across examples in which the modifier can take a plural marker, which is not always justifiable in terms of the meaning of a compound, e.g. *products sector – product sector, points system – point system*. These are opposed to examples in which the plural marker on the modifier seems to be important for meaning construal, e.g. *girls fan – girl fan, farms land – farmland, islands area – island area*, etc. The analysis of the use of the plural marker in the absence of grammatical requirements may help understand the reasons for this phenomenon.

It was also observed that some semantic relations occur more often than the others, for example, almost two thirds of all the compounds in the collected data set realised only three semantic relations: PURPOSE (34%), POSSESSION (16.2%) and LOCATION (14.4%). The reasons for the preferential use of these three semantic relations may be the result of the relative broadness of these categories. At the same time, it might be the case that there are other reasons that explain it. For example, it may be suggested that these three categories denote basic relational concepts that connect objects and phenomena of extralinguistic reality. A closer look into this issue may be beneficial for understanding the process of inferring the relational information to connect items of language into sensible sequences.

These and a number of other points briefly pointed out in the course of this thesis demonstrate the richness and complexity of the topic and there are a number of challenges involved in the research. I view the results obtained in the course of my study as an opening of the opportunities for further research in this field.

Reference list

Ackema, P., & Neelman, A. 2004. *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: Oxford University Press.

Albright, A., & Hayes, B. 2003. Rules vs. analogy in English past tenses: A computational / experimental study. *Cognition*, 90(2). 119-161.

Allan, K. 2009. Stereotype semantics. In K. Allan (ed.), *Concise Encyclopaedia of Semantics*, 939-942. Oxford, Amsterdam: Elsevier.

Allen, M.R. 1978. *Morphological Investigations.* University of Connecticut, Connecticut: Storrs.

Allen, M.R. 1980. Semantic and phonological consequences of boundaries: A morphological analysis of compounds. In: M. Aronoff, & M.L. Kean (eds.), *Juncture*, 9-27. Saratoga, CA: Amma Libri.

Anderson, J.M. 1971. *The Grammar of Case: Towards a Localistic Theory*. London: Cambridge University Press.

Anderson, S. R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Andrews, S., Miller, B., & Rayner, K. 2004. Eye fixation measures of morphological segmentation of compound words: There is a mouse in the mousetrap. *European Journal of Cognitive Psychology*, 16(2). 285-311.

Apresyan, Y.D. 1974. (Апресян, Ю.Д. *Лексическая семантика: Синонимические средства языка.* – М.: Наука.)

Arndt-Lappe, S. 2011. Towards an exemplar-based model of stress in English noun-noun compounds. *Journal of Linguistics*, 47(3). 549-585.

Arndt-Lappe, S. 2012. *Regularity and Variability in English word-formation: A case study.* Ms. Universität Siegen.

Arndt-Lappe, S., & Bell, M. 2012. *Explaining compound stress analogically*. Paper presented at the 15th International Morphology Meeting, Vienna, February 9-12.

Arnold, I.V. 1973. *Lexicology of the Modern English Language*. Moscow: Publishing House of Literature in Foreign Languages.

Aronoff, M. 1976. *Word Formation in Generative Grammar. Linguistic Inquiry Monograph 1*. Cambridge, MA: MIT Press.

Baayen, R.H. 1992. Quantitative aspects of morphological productivity. In G. Booij, & J. van Marle (eds.), *Yearbook of Morphology 1991*, 109-149. Dordrecht: Kluwer.

Baayen, R.H. 1993. Frequency, transparency and productivity. In G. Booij, & J. van Marle (eds.), *Yearbook of Morphology 1992*, 181-208. Dordrecht: Kluwer.

Baayen, R.H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R.H. 2010. The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (ed.), *New Impulses in Word-Formation (Linguistische Berichte Sonderheft 17)*, 383-402. Hamburg: Buske.

Baayen, R.H., & Lieber, R. 1991. Productivity and English derivation: A corpus based study. *Linguistics,* 29(5). 801-843.

Baayen, R. H., Dijkstra, T., & Schreuder, R. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory & Language*, 37(1). 94-117.

Bates, D. M., Maechler, M., & Dai, B. 2007. *lme4: Linear mixed-effects models using S4 classes.* Retrieved from http://lme4.r-forge.r-project.org/ on 23 July 2012.

Bauer, L. 1978. *The Grammar of Nominal Compounding with Special Reference to Danish, English and French.* Odense: Odense University Press.

Bauer, L. 1979. On the need for pragmatics in the study of nominal compounding. *Journal of Pragmatics,* 31(1). 45-50.

Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Bauer, L. 1990. Be-heading the word. *Journal of Linguistics,* 26(1). 1-31.

Bauer, L. 1998. When is a sequence of two nouns a compound in English? *English Language & Linguistics*, 21(1). 65-86.

Bauer, L. 2001. *Morphological Productivity*. Cambridge; New York: Cambridge University Press.

Bauer, L. 2006. Compounds and minor word-formation types. In B. Aarts, & A. Mc-Mahon (eds.), *Handbook of English* Linguistics, 483-506. Oxford: Basil Blackwell.

Bauer, L., & Huddleston, R. 2002. Lexical word-formation. In R. Huddleston, & G. K. Pullum (eds), *The Cambridge Grammar of the English Language*, 1621-1721. Cambridge: Cambridge University Press.

Bauer, L., & Tarasova, E. (to appear). The meaning link in nominal compounds.

Bauer, L., & Valera, S. 2005. Conversion or zero-derivation: An introduction. In L. Bauer, & S. Valera (eds.), *Approaches to Conversion/Zero-Derivation,* 7-17. Munster: Waxmann.

Bauer, L., Lieber, R., & Plag, I. (to appear). *English Morphology: A Reference Guide to Contemporary English Word-Formation and Inflection.* Oxford: Oxford University Press.

Bell, M.J. 2005. Against nouns as syntactic premodifiers in English noun phrases. *Working Papers in English and Applied Linguistics*, 11(1). 1-48.

Bell, M.J. & Plag, I. 2012. Informativeness is a determinant of compound stress in English. *Journal of Linguistics*, 48(3). 485-520. Available on CJO 2012 doi:10.1017/S0022226712000199.

Bell, M.J. 2011. At the boundary of morphology and syntax: NN-constructions in English. In A. Galani, G. Hicks & G. Tsoulos (eds.), *Morphology and its Interfaces*, 137-168. Amsterdam and Philadelphia: John Benjamins.

Benczes, R. 2006a. *Creative Compounding in English: The Semantics of Metaphorical and Metonymical Noun-Noun Combinations.* Amsterdam; Philadelphia: Benjamins.

Benczes, R. 2006b. Analysing metonymical noun–noun compounds: The case of *freedom fries*. In R. Benczes & S. Csábi (eds.), *The Metaphors of Sixty: Papers Presented on the Occasion of the 60th Birthday of Zoltán Kövecses,* 46-54. Budapest: Eötvös Loránd University.

Benczes, R. 2009. What motivates the production and use of metaphorical and metonymical compounds? In M. Brdar, M. Omazić and V. P. Takač (eds.), *Cognitive Approaches to English. Fundamental, Methodological, Interdisciplinary and Applied Aspects*, 49-67. Newcastle-upon-Tyne: Cambridge Scholars.

Berman, R. 1986. The Acquisition of Morphology/Syntax: A Crosslinguistic Perspective. In P. Fletcher & M. Garman (eds.), *Language Acquisition*, *2nd edition*, 429-447. Cambridge: Cambridge University Press.

Berman, R. 2009. Children's acquisition of compound constructions. In R. Lieber & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 298-322. Oxford: Oxford University Press.

Berman, R.A., & Ravid, D. 1986. Lexicalization of Noun Compounds. *Hebrew Linguistics,* 24(1). 5-22. [in Hebrew]

Berry, M. 1975*. Introduction to Systemic Linguistic I: Structure and Systems*. London: B.T. Batsford Ltd.

Bertram, R., & Hyönä, J. 2003. The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, 48(3). 615-634.

Besten, H. den. 1983. On the interaction of root transformations and lexical deletive rules. In W. Abraham (ed.), *On the Formal Syntax of the Westgermania*, 47-131. Amsterdam; Philadelphia: Benjamins.

Biber, D., Finegan E., Conrad S., et. al. 2000. *Longman Grammar of Spoken and Written English*. New York: Pearson Ed. Limited.

Bien, H., Levelt, W. J., & Baayen, R. H. 2005. Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America*, 102. 17876-17881.

Bisetto, A., & Scalise, S. 2005. The classification of compounds. *Lingue e linguaggio*, 4(2). 319-332.

Bisetto, A., & Scalise, S. 2009. Classification of compounds. In R. Lieber, & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 49-82. Oxford: Oxford University Press.

Blevins, J. P., & Blevins, J. 2009. Introduction: Analogy in Grammar. In J. P. Blevins, & J. Blevins (eds.), *Analogy in Grammar: Form and Acquisition*, 1-12. Oxford: Oxford University Press.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge: MIT Press.

Bloomfield, L. 1933. *Language*. New York: Holt.

Boldyrev, N.N. 2000. (Болдырев, Н.Н. *Когнитивная семантика: Курс лекций по английской филологии*. – Тамбов: ТГУ.)

Booij, G. 1977. *Dutch Morphology: a Study of Word Formation in Generative Grammar.* Lisse: The Peter De Ridder Press.

Booij, G. 2005. Compounding and derivation: Evidence for Construction Morphology. In W.U. Dressler, D. Kastovsky, O.E. Pfeiffer, & F. Rainer (eds.), *Morphology and its Demarcations:*

Selected Papers from the 11th Morphology Meeting, Vienna, February 2004, 109-132. Amsterdam; Philadelphia: Benjamins.

Booij, G. 2007. Construction morphology and the lexicon. In F. Montermini, G. Boyé, & N. Hathout (eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, 34-44. Somerville, MA: Cascadilla Proceedings Project.

Booij, G. 2010. Compound construction: Schemas or analogy? A Construction Morphology perspective. In S. Scalise, & I. Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 93-108. Amsterdam; Philadelphia: Benjamins.

Booij, G. 2011. Morphology in Construction Grammar. In T. Hoffman, & G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Boroditsky, L. 2000. Metaphoric Structuring: Understanding time through spatial metaphors. *Cognition*, 751(1). 1-28.

Bradley, D. 1980. Lexical representation of derivational relation. In: M. Aronoff, & M.L. Kean (eds.), *Juncture*, 37-55. Saratoga, CA: Amma Libri.

Brekle, H.E. 1976. *Generative Satzsemantik im System der englischen Nominalkomposition*. München: Fink.

Brinton, L. & Traugott, E.C. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.

Brugmann, K. 1891. *A comparative grammar of the Indo-Germanic languages: A Concise Exposition of the History of Sanskrit, Old Iranian (Avestin and Old Persian), Old Armenian, Greek, Latin, Umbro-Samnitic, Old Irish, Gothic, Old High German, Lithuanian and Old Church Slavonic.* Translated from German by R.S. Conway, & W.H.D. Rouse. Vol.2, pt 2. New York: B. Westermann.

Buchsbaum, B.R., Hickok, G., & Humphries, C. 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25(5). 663-678.

Bundgaard, P.F., Østergaard S., & Stjernfelt, F. 2006. Waterproof fire stations? *Semiotica*, 161(1-4). 363-393.

Butterworth, B. 1983. Lexical representation. In B. Butterworth (ed.), *Language Production, vol. 2 Development, Writing and Other Language Processes*, 257-294. London: Academic Press.

Bybee, J. 1995. Regular morphology and the lexicon. *Language & Cognitive Processes*, 10(5). 425-455.

Cambridge Online. 2011. Retrieved from http://dictionary.cambridge.org/

Carstairs-McCarthy, A. 2010. Review of Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 2009. Oxford: Oxford University Press. *Word Structure*, 3(2). 252-282.

Chapman, D., & Skousen, R. 2005. Analogical modeling and morphological change: The case of the adjectival negative prefix in English. *English Language and Linguistics*, 9(2). 333-357.

Chomsky, N. 1970. Remarks on Nominalization. In R. Jacobs, & P. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184-221. Waltham, MA: Ginn.

Cienki, A. 2007. Frames, idealized cognitive models, and domains. In D. Geeraerts & H. Cuyckens (eds.), *Handbook of Cognitive Linguistics,* 170-187. Oxford: Oxford University Press.

Clark, E. V. 1993. *The Lexicon in Acquisition.* Cambridge: Cambridge University Press.

Clark, H.H. 1973. Space, time, semantics and the child. In T.E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, 27-63. New York: Academic Press.

Clark, E. V., Hecht, B. F., & Mulford, R. C. 1986. Coining complex compounds in English: Affixes and word order in acquisition. *Linguistics*, 24(1). 7-29.

Collins Online. 2012. Retrieved from http://www.collinsdictionary.com/

Connolly, A. C., Fodor, J. A., Gleitman, L. R., & Gleitman, H. 2007. Why stereotypes don't even make good defaults. *Cognition,* 103(1). 1-22.

Connor, C. 2010. *A Diachronic Exploration of the Contribution of the Harvesting of the Marine Environment to a Distinctive New Zealand English Lexicon*. Unpublished Ph.D dissertation, Victoria University of Wellington.

Coulson, S. 2001. *Semantic Leaps: Frame-shifting and Conceptual Blending in Meaning Construction.* New York and Cambridge: Cambridge University Press.

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D.A. 2006. *A Glossary of Semantics and Pragmatics*. Edinburgh: Edinburgh University Press.

Cruse, D.A., & Croft, W. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Davies, M. 2004- *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

Davies, M. 2008- *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

De Vaan, L., Schreuder, R., & Baayen, R.H. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *Mental Lexicon*, 2(1). 1-23.

Derwing, B.I., & Skousen, R. 1994. Productivity and the English past tense: Testing Skousen's analogical model. In S.D. Lima, R. Corrigan, & G.K. Iverson (eds.), *The Reality of Linguistic Rules*, 193-218. Amsterdam & Philadelphia: John Benjamins.

Di Sciullo, A. M., & Williams, E. 1987. *On the Definition of Word*. Cambridge: The MIT Press.

Dillon J. 2000. *The Question of Being*. In J. Brunschwig & G.E.R. Lloyd (eds.), *Greek thought. A guide to classical knowledge,* 51-71. Harvard: Harvard University Press.

Downing, P. 1977. On the creation and use of English compound nouns. *Language*, 53(4). 810-842.

Dressler, W. U. 2005. Word-formation in Natural Morphology. In P. Stekauer & R. Lieber (eds.), *Handbook of Word-Formation (Studies in Natural Language and Linguistics Theory, volume 64)*, 335-352. Dordrecht: Springer.

Elzinga, D. 2006. English adjective comparison and analogy. *Lingua*, 116(6). 757-770.

Estes, Z. 2003. Attributive and relational processes in nominal combination. *Journal of Memory & Language*, 48(2). 304-319.

Estes, Z. 2003. Attributive and relational processes in nominal combination. *Journal of Memory and Language*, 48(3). 304-319.

Estes, Z,. & Jones, L. 2006. Priming via relational similarity: A COPPER HORSE is faster when seen through a GLASS EYE. *Journal of Memory & Language*, 55(1). 89-101.

Estes, Z., & Jones, L. 2008. Relational processing in conceptual combination and analogy. *Behavioral & Brain Sciences,* 31(4). 385-386.

Estes, Z., Golonka, S., & Jones, L. L. (to appear). Thematic thinking: The apprehension and consequences of thematic relations. *Psychology of Learning and Motivation.*

Evans V. 2009. *How Words Mean*. New York: Oxford University Press.

Fabb, N. 1998. Compounding. In A. Spencer and A.M. Zwicky (eds.), *The Handbook of Morphology*, 66-83. Oxford: Blackwell.

Fauconnier, G. 1997. *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fauconnier, G. 2007. Mental spaces. In D. Geeraerts & H. Cuyckens (eds.), *Handbook of Cognitive Linguistics,* 351-376. Oxford: Oxford University Press.

Fauconnier, G., & Turner, M. 1995. *Conceptual integration and formal expression*. Available at http://zakros.ucsd.edu/trohrer/metaphor/turner.htm (11.03.2013)

Fauconnier, G., & Turner, M. 1998. Conceptual integration networks. *Cognitive Science,* 22(2). 133-87.

Fauconnier, G., & Turner, M. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fauconnier, G., & Turner, M. 2003. Conceptual Blending, form and meaning. *Recherches en communication: Sémiotique Cognitive*, 19(1), 57-86.

Ferris, C. 1993. *The Meaning of Syntax: a Study of Adjectives in English.* London & New York: Longman.

Fillmore, C. 1972. *Essentials of English Grammar.* New York: Holt, Rinehart & Winston.

Fodor, J. A. 1998. *Concepts: where cognitive science went wrong*. Oxford: Clarendon.

Fodor, J. A., & Lepore, E. 1996. The red herring and the pet fish: Why concepts can't be prototypes. *Cognition,* 58(2). 253-270.

Frisson S., Niswander-Klement E., & Pollatsek, A. 2008. The role of semantic transparency in processing of English compound words. *British Journal of Psychology*, 99(1). 87-107.

Fusté-Herrmann, B. 2008. *Idiom Comprehension in Bilingual and Monolingual Adolescents*. Unpublished PhD Thesis. University of South Florida.

Gagné, C.L. 2000. Relation-based combinations versus property based combinations: A test of the CARIN theory and dual process theory of conceptual combination. *Journal of Memory & Language,* 42(3). 365-389.

Gagné, C.L. 2001. Relational and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(1). 236-254.

Gagné, C.L., & Shoben, E.J. 1997. The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23(1). 71-87.

Gagné, C.L., & Shoben, E.J. 2002. Priming relations in ambiguous noun-noun combinations. *Memory & Cognition*, 30(4). 637-646.

Gagné, C.L., & Spalding, T.L. 2004. Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain & Language*, 90(1-3). 478-486.

Gagné, C.L., & Spalding, T.L. 2006a. Conceptual combination: Implications for the mental lexicon. In G. Libben, & G. Jarema (eds.), *The Representation and Processing of Compound Words*, 145-168. New York: Oxford University Press.

Gagné, C.L., & Spalding, T.L. 2006b. Using conceptual combination research to better understand novel compound words. *SKASE Journal of Theoretical Linguistics*, *3.* 9-16.

Gagné, C.L., & Spalding, T.L. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory & Language,* 60(1). 20-35.

Gagné, C.L., & Spalding, T.L. 2010. Relational competition during compound interpretation. In S. Scalise, & I. Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 287-300. Amsterdam; Philadelphia: Benjamins.

Geeraerts, D., & Cuyckens, H. 2007. Introducing cognitive linguistics. In D. Geeraerts, & H. Cuyckens (eds.), *Oxford Handbook of Cognitive Linguistics*, 1-21. Oxford; New York: Oxford University Press.

Giegerich, H.J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language & Linguistics*, 81(1). 1-24.

Giegerich, H.J. 2005. Associative adjectives in English and the lexicon-syntax interface. *Journal of Linguistics*, 41(3). 571-591.

Giegerich, H.J. 2009. Compounding and Lexicalism. In R. Lieber & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 178-200. Oxford: Oxford University Press.

Gleitman, L.R., & Gleitman, H. 1970. *Phrase and Paraphrase: Some Innovative Uses of Language.* New York: W.W. Norton.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Granville Hatcher, A. 1960. An introduction to the analysis of English noun compounds. *Word* 16(3), 356-373.

Grzega, J. 2003. Borrowing as a word-finding process in Cognitive Historical Onomasiology. *Onomasiology Online* 4, 22-42.

Guevara, E., & Scalise, S. 2009. Searching for universals in compounding. In S. Scalise, E. Magni, & A. Bisetto (eds.) *Universals of Language Today*, 101-128. Berlin: Springer.

Halle, M. 1973. Prolegomena to a theory of word-formation. *Linguistic Inquiry*, 41(1). 3-16.

Halliday, M.A.K. 1976. *Halliday: System and Function in Language. Selected papers edited by G.R. Kress.* London: Oxford University Press.

Halliday, M.A.K. 2002. *On Grammar. Collected Works of M.A.K. Halliday.* J. Webster (ed.). London: Continuum.

Hamburger, H.G., & Jacobsen, A. 1972. *Ny Ord i Dansk 1968-1969*. København: Danske Sprognævn.

Hampton, J.A. 1987. Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, 151(1). 55-71.

Hay, J., & Baayen, R.H. 2005. Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9(7). 342-348.

Heyvaert, L. 2009. Compounding in Cognitive Linguistics. In R. Lieber & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 233-254. Oxford: Oxford University Press.

Hickok, G., Houde, J., & Rong, F. 2011. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3). 407-422.

Jaarsveld, H.J. van, Coolen, R., & Schreuder, R. 1994. The role of analogy in the interpretation of novel compounds. *Journal of Psycholinguistic Research,* 23(1). 111-137.

Jaarsveld, H. van, & Rattink, G. 1988. Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, 17(6). 447-473.

Jackendoff, R. 1975. Morphological and semantic regularities in the lexicon.*Language*, 51(3). 639-671.

Jackendoff, R.S. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., Libben, G. 1999. Processing compounds: A cross-linguistic study. *Brain and Language,* 68(1-2). 362-369.

Jespersen, O. 1942. *A Modern English Grammar on Historical Principles. Pt 6. Morphology*. London: Allen & Unwin.

Juhasz, B.J., Inhoff, A.W., & Rayner, K. 2005. The role of interword spaces in the processing of English compounds. *Language & Cognitive Processes,* 20(2). 291-316.

Kavka, S. 2009. Compounding and idiomatology. In R. Lieber, & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 19-33. Oxford: Oxford University Press.

Kemmerer, D. 2005. The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia*, 43(5). 797-806.

Keuleers, E. 2008. *Memory-based learning of inflectional morphology*. Antwerpen: University of Antwerpen dissertation.

Koziol, H. 1937. *Handbuch der Englische Wortbildungslehre*. Heidelberg: Winter.

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

Krott, A., Schreuder, R., Baayen, R. H., & Dressler, W. U. 2007. Analogical effects on linking elements in German compound words. *Language & Cognitive Processes,* 22(1). 25-57.

Kruisinga, E., & Erades, P.A. 1953. *An English Grammar. Vol.1. Accidence and Syntax*. Groningen: Noordhoff.

Kunter, G. 2011. *Compound Stress in English. The Phonetics and Phonology of Prosodic Prominence*. Berlin: de Gruyter.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: CSLI.

Lakoff, G. 1993. The contemporary theory of metaphor. In A. Ortony (ed.), *Metaphor and Thought*. 2nd ed, 202-251. Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought.* New York: Basic Books.

Langacker, R.W. 1987a. *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prequisities.* Stanford: Stanford University Press.

Langacker, R.W. 1987b. *Foundations of Cognitive Grammar. Vol. 2: Descriptive Application.* Stanford: Stanford University Press.

Langacker, R.W. 1990. *Concept, Image, and Symbol: The Cognitive Basis of Grammar.* Berlin; New York: de Gruyter.

Langacker, R.W. 1999. *Grammar and conceptualization.* Berlin: De Gruyter.

Langacker, R.W. 2000. A dynamic usage-based model. In S. Kemmer & M. Barlow (eds), *Usage Based Models of Language*, 1-63. Stanford: CSLI.

Langacker, R.W. 2003. Constructional integration, grammaticization, and serial verb constructions. *Language and Linguistics*, 4(2). 251-278.

Laudana, A., & Burani, C. 1995. Distributed properties of derivational affixes: Implications for processing. In L.B Feldman (ed.), *Morphological Aspects of Language Processing: Cross-Linguistic Perspectives,* 345-364. Hillsdale, NJ: Erlbaum.

Leech, R., Mareschal, D., & Cooper, R. P. 2008. Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral & Brain Sciences*, 31(4). 357-378.

Lees, R.B. 1968. *The Grammar of English Nominalizations*. The Hague: de Gruyter.

Lees, R.B. 1970. On very deep grammatical structure. In R.A. Jacobs, & P.S. Rosenbaum (eds.), *Readings in English Transformational Grammar*. Waltham; Massachusetts: Ginn.

Levelt, W.J.M. 2001. Spoken word production: A theory of lexical access. *PNAS. Proceedings of the National Academy of Sciences,* 98(23). 13464-13471.

Levelt, W.J.M., Roelofs, A.P.A. & Meyer, A.S. 1999. A theory of lexical access in speech production. *Behavioral & Brain Sciences,* 22(1). 1-37.

Levi, J. 1978. *The Syntax and Semantics of Complex Nominals.* New York: Academic Press Inc.

Levin, B. 1993. *English Verb Classes and Alterations*. Chicago: University of Chicago Press.

Libben, G. 1998. Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain & Language*, 61(1). 30-44.

Libben, G. 2005. Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics*, 50(1-4). 267-283.

Libben, G. 2006. Why study compound processing? An overview of the issues. In G. Libben, & G. Jarema (eds.), *Compound Processing*, 1-22. New York: Oxford University Press.

Libben, G. & Jarema, G. 2006. *The Representation and Processing of Compound Nouns*. New York, Oxford University Press Inc.

Libben, G., Gibson, M., Yoon, Y-B., & Sandra, D. 2003. Compound fracture: The role of semantic transparency and morphological headedness. *Brain & Language,* 84(1). 26-43.

Lieber, R. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.

Lieber, R. 2010. On the lexical semantics of compounds: Non-affixal (de)verbal compounds. In S. Scalise, & I. Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 127-144. Amsterdam; Philadelphia: Benjamins.

Lieber, R., & Štekauer, P. 2009. Introduction: Status and definition of compounding. In R. Lieber & P. Štekauer (eds.), *The Oxford Handbook of Compounding*, 3-18. Oxford: Oxford University Press.

Likhachev, D.C. 1997. (Лихачев, Д.С. Концептосфера русского языка // *Русская словесность*. – М.: Academia, 1997. – С. 280-288.)

Lindner, S. 1981. *A Lexico-Semantic Analysis of English Verb-Particle Constructions with UP and OUT*. Unpublished doctoral dissertation. San Diego, Amsterdam and Philadelphia.

Lipka, L. 1994. Lexicalization and institutionalization. In R.E. Asher (ed.), *The Encyclopedia of Language and Linguistics.* 1st ed., 2164-2167. Oxford: Pergamon Press.

Longman Online. 2012. Retrieved from http://www.longmandictionariesonline.com/

Lyons, J. 1978. *Semantics*. Vol. 2. Cambridge, New York: Cambridge University Press.

Maguire, P., Maguire, R., & Cater, A. W. S. 2010. The influence of interactional semantic patterns on the interpretation of noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3). 288-297.

Maguire, P., Wisniewski, E.J., & Storms, G. 2010. A corpus study of semantic patterns in compounding. *Corpus Linguistics & Linguistic Theory,* 6(1). 49-73.

Marchand, H. 1960. *The Categories and Types of Present-Day English Word-Formation; A Synchronic-Diachronic Approach, by Hans Marchand.* Chicago: University of Alabama Press.

Marchand, H. 1965. On the analysis of substantive compounds and suffixal derivatives not containing a verbal element. *Indogermanische Forschungen*, 70(1). 117-145.

Marchand, H. 1969. *The Categories and Types of Present-Day English Word Formation: A Synchronic-Diachronic Approach.* 2nd edn. Munchen: C. H. Beck'sche Verlagsbuchhandlung.

Marchand, H. 1974. *Studies in Syntax and Word Formation.* Edited by D. Kastovsky. Internationale Bibliothek für Allgemeine Linguistik, Bd. 18. Munich: Fink.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review*, 101(1). 3-33.

Martinovski, B. 1995. Three cognitive theories: major differences and similarities: Comparing the cognitive theories of language of Jackendoff, Langacker and Talmy. *http://www.ling.gu.se/~biljana/st1-97/cogn.html*

Matlock, T., Ramscar, M., & Boroditsky, L. 2005. On the experiential link between spatial and temporal language. *Cognitive Science*, 29(4). 655-664.

Matthews, P.H. 1974. *Morphology: An Introduction to the Theory of Word-Structure.* Cambridge: Cambridge University Press.

Medin, D.L., & Shoben, E.J. 1988. Context and structure in conceptual combination. *Cognitive Psychology*, 20(1). 158-190.

Michaelis, L.A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics*, 15(1). 1-67.

Murphy, G.L. 1988. Comprehending complex concepts. *Cognitive Science*, 12(4). 529-562.

Nakov, P.I. 2007. *Using the Web as an Implicit Training Set: Application to Noun Compound Syntax and Semantics*. Unpublished PhD Thesis. University of California at Berkeley.

Ó Séaghdha, D. 2008. *Learning Compound Noun Semantics. Technical report.* Cambridge: University of Cambridge Computer Laboratory.

OED Online. 2012. Retrieved from http://www.oed.com/

Olsen, S. 2000. Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte*, 181(1). 55-70.

Ornstein, R. 1969. *The Experience of Time*. Baltimore, MD: Penguin Books.

Oxford Online. 2012. Retrieved from http://oxforddictionaries.com/

PASW Statistics software, version 18.0.3 - September 2010.

Pelletier, F.J. 1994. The Principle of Semantic Compositionality. *Topoi*, 13(1). 11-24.

Pinheiro, J. C., & Bates, D. M. 2000. *Mixed Effects Models in S and S-Plus*. New York: Springer.

Plag, I. 2003. *Word-Formation in English*. New York: Cambridge University Press.

Plag, I. 2006. The variability of compound stress in English: Structural, semantic and analogical factors. *English Language & Linguistics*, 10(1). 143-172.

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

Plag, I., Kunter, G., & Lappe, S. 2007. Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics & Linguistic Theory*, 3(2). 199-233.

Plag, I., Kunter, G., Lappe, S., & Braun, M. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language*, 84(4). 760-794.

Plungyan, V.A., & Rakhilina, E.V. 1998. (Плунгян В.А., Рахилина Е.В. *Парадоксы валентностей.* – Семиотика и информатика, вып. 36). Available at http://lpcs.math.msu.su/~uspensky/journals/siio/36/36PLURAKH.pdf (19.03.2013)

Pollatsek A., & Hyönä J. 2006. Processing of morphologically complex words in context: What can be learned from eye movements. In S. Andrews (ed.), *From Inkmarks to Ideas: Current Issues in Lexical Processing*, 275-298. Hove: Psychology Press.

Pollatsek, A., Hyönä, J., & Bertram, R. 2000. The role of constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance,* 26(2). 820-833.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.

R Development Core Team 2011. Vienna: R Foundation for Statistical Computing.

Radden, G. 2004. The metaphor TIME AS SPACE across languages. In N. Baumarten, C. Böttger, M. Motz, & J. Probst (eds.), *Übersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung – das Leben mit mehreren Sprachen: Festschrift für Juliane House zum 60. Geburtstag,* 225-238. Bochum: AKS-Verlag.

Radden, G., & Dirven, R. 2007. *Cognitive English Grammar*. Amsterdam; Phildelphia: Benjamins.

Raffray, C. N., Pickering, M. J., & Branigan, H. P. 2007. Priming the interpretation of noun-noun combinations. *Journal of Memory & Language,* 57(3). 380-395.

Ricca, D. 2010. Corpus data and theoretical implications: With special reference to Italian V-N compounds. In S. Scalise, & I. Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 237-254. Amsterdam: Benjamins.

Ryder, M.E. 1994. *Ordered Chaos: The Interpretation of English Noun-Noun Compounds*. Berkeley: University of California Press.

Sandra, D. 1994. The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language & Cognitive Processes*, 9(2). 227-269.

Sandra, D., & Rice, S. 1995. Network analyses of prepositional meaning: Mirroring whose mind — the linguist's or the language users? *Cognitive Linguistics*, 6(1). 89-130.

Scalise, S. 1986. *Generative Morphology*. Dodrecht: Foris.

Scalise, S., & Guevara, E. 2005. The Lexicalist approach to word-formation and the notion of the lexicon. In P. Štekauer, & R. Lieber (eds.), *Handbook of Word-Formation*, 147-187. Dordrecht: Kluwer.

Scalise, S., & Vogel, I. 2010. Why compounding? In S. Scalise & I. Vogel (eds.), *Cross-Disciplinary Issues in Compounding*, 1-18. Amsterdam; Philadelphia: Benjamins.

Scheler, M. 1977. *Der Englische Wortschatz*. Berlin: Schmidt.

Schlücker, B., & Plag, I. 2011. Compound or phrase? Analogy in naming. *Lingua,* 121(9). 1539-1551.

Schmid, H-J. 2007. Entrenchment, salience and basic levels. In D. Geeraerts & H. Cuyckens (eds.), *Handbook of Cognitive Linguistics,* 169-203. Oxford: Oxford University Press.

Schreuder, R., & Baayen, R.H. 1995. *Modelling Morphological Processes.* Hillsdale, NJ, England: Erlbaum.

Schreuder, R., & Baayen, R.H. 1997. How complex simple words can be. *Journal of Memory & Language*, 37(1). 118-139.

Selkirk, E.O. 1982. *The syntax of words*. (Linguistic Inquiry Monographs 7.) Cambridge, MA: MIT Press.

Siegel, M.E.A. 1980. *Capturing the Adjective.* New York; London: Garland.

Skousen, R. 1989. *Analogical Modeling of Language.* Dordrecht: Kluwer.

Smirnitsky, A.I. 1956. (Смирницкий, А.И. *Лексикология английского языка.* – М.: Изд-во лит-ры на ин. языках.)

Spellman, B. A., Holyoak, K. J., & Morrison, R. G. 2001. Analogical priming via semantic relations. *Memory & Cognition*, 29(3). 383-393.

Spencer, A. 1997. On the redundancy of lexical categories. Paper presented at *Lexical Functional Grammar'97*, San Diego, June 1997. Available at http://privatewww.essex.ac.uk/~spena/papers/redundcats.pdf (18.02.2013)

Sperber, D. & Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Štekauer, P. 2000. *English Word-Formation. A History of Research 1960-1995.* Tübingen: Narr.

Štekauer, P. 2005. *Meaning Predictability in Word Formation*. Amsterdam; Philadelphia: Benjamins.

Stepanov, Y.S. 2007. (Степанов, Ю.С. *Концепты. Тонкая пленка цивилизации*. – М.: Языки славянских культур.)

Storms, G., & Wisniewski, E.J. 2005. Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition*, 33(5). 852-861.

Sweet, H. 1900. *A New English Grammar Logical and Historical. Pt 1.* Oxford: Clarendon Press.

Sweetser, E. 1999. Compositionality and blending: Semantic composition in a cognitively realistic framework. In G. Redeker, & T. Janssen (eds.), *Cognitive Linguistics: Foundations, Scope, and Methodology*, 129-162. Berlin; New York: de Gruyter.

Szymanek, B. 1988. *Categories and Categorization in Morphology*. Lublin: Katolicki Uniwersytet Lubelski.

Taft, M. 1981. Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior*, 20(3). 289-297.

Taft, M. 2003. Morphological representations: A correlation between form and meaning. In E.M.H. Assink, & D.Sandra (eds.), *Reading Complex Words: Cross-Language Studies*, 113-137. Amsterdam: Kluwer Academic.

Talmy, L. 2000. *Toward a Cognitive Semantics. Vol. 1 Concept Structuring Systems*. Cambridge, MA: MIT Press.

Tarasova, E.I. 2007. (Тарасова, Е. И. *Номинативный бином в свете проблемы концептуальной целостности при вербальной раздельнооформленности*: Диссертация … филологических наук. 10.02.19. – Тверь.)

Taylor, J.R. 2002. *Cognitive Grammar.* Oxford: Oxford University Press.

Taylor, J.R. 2008. Prototypes in cognitive linguistics. In *P. Robinson, & N.C. Ellis (eds.), Handbook of Cognitive Linguistics and Second Language Acquisition*, 39-65. New York, London: Routledge.

Turner, M. 1996. *The Literary Mind: The Origins of Thought and Language*. Chicago; London: University of Chicago Press.

Ungerer, F. 2007. Word-Formation. In D. Geeraerts, & A. Cuyckens (eds.), *Handbook of Cognitive Linguistics*, 650-675. Oxford: Oxford University Press.

Volodarskaya, E.F. 2002. (Володарская, Э.Ф. Заимствование как отражение русско-английских контактов // *Вопросы языкознания*. –№4, – 96-118.)

Warren, B. 1978. *Semantic Patterns of Noun-Noun Compounds.* Goethenburg: Acta Universitatis Gothoburgensis.

Webster Online. 2012. Retrieved from http://www.merriam-webster.com/

Weiskopf, D. 2007. Compound nominals, context, and compositionality. *Synthese*, 156(1). 161-204.

Wierzbicka A. 1980. *The Case for Surface Case*. Ann Arbor, MI: Karoma.

Williams, E. 1981. On the notions 'Lexically Related' and 'Head of a Word'. *Linguistic Inquiry*, 12(3). 245-274.

Windsor, J. 1993. The functions of novel word compounds. *Journal of Child Language*, 20(1). 119-138.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E. & Frackowiak, R. 1990. Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, 114(4). 1803-1817.

Wisniewski, E.J. 1997. When concepts combine. *Psychonomic Bulletin & Review*, 4(2). 167-183.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. 2008. *Formulaic Language: Pushing the Boundaries*. Oxford UK: Oxford University Press.

Zimmer, K.E. 1971. Some general observations about nominal compounds. *Working Papers on Language Universals*, 5(1). 1-21.

Zwitserlood, P. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language & Cognitive Processes*, 9(2). 341-368.

Zwitserlood, P., Bolwiender, A., & Drews, E. 2005. Priming morphologically complex verbs by sentence contexts: Effects of semantic transparency and ambiguity. *Language & Cognitive Processes,* 20(3). 395-415.

List of appendices[112]

Appendix 1. The list of 500 compounds initially collected for the analysis

Appendix 2. The list of 50 compounds whose constituents were used for further data collection

Appendix 3. Results obtained from the raters

Appendix 4. Collected constituent word families (CWFs)

Appendix 5. Coded data table

Appendix 6. Grouping of compound constituents according to the degree of positional preference

[112] See the attached CD.