

NON-CANONICAL COMPOUNDS IN THE DISCOURSE OF ELECTRICAL ENGINEERING

Miloš D. Đurić

Faculty of Electrical Engineering, University of Belgrade, Serbia
Phone: +381113218366, E-Mail: djuric@etf.bg.ac.rs, djuric@etf.rs

Abstract. *Taking into account some general definitions of compounds found in the pertinent linguistic literature and their frequent syntactic features (such as fixed order and recursiveness) we have analysed and investigated semantic compositionality and some other features pertaining to these linguistic items. Furthermore, we have drawn attention to the possibility of implementing certain criteria in connection with the distribution of non-canonical compound constituents. It has been pointed out that distributional evidence is applicable for English non-canonical compounds in electrical engineering discourse. Moreover, distribution is used as an indicator of lexical integrity of these linguistic items. Apart from fairly clear cases, different examples of English non-canonical compounds taken from the specialised discourse of electrical engineering have been examined by way of illustration and certain results of our analysis have been presented.*

Key words: *Non-Canonical Compounds, the English Language, Discourse of Electrical Engineering, Fixed Order, Recursiveness, Distributional Evidence, Semantic Compositionality.*

1. INTRODUCTORY REMARKS

Generally speaking, it seems to us that English non-canonical compounds (i.e. multi-constituent compounds) have been studied less than canonical compounds (i.e. two-constituent/binary compounds) in the pertinent linguistic literature. Even when these non-canonical compounds are discussed, it mainly occurs as a sort of brief mentioning after elaborate description of canonical compounds¹. As a consequence of such traditional approaches, non-canonical compounds have been imprecisely determined, fuzzily established and have remained an unstable category. Moreover, these items have not been studied so often in the specific discourses. Since English manifests a wealth in the number and variety of non-canonical compounds, primarily in specific discourses, we have decided to analyse non-canonical compounds in the discourse of electrical engineering.

There is no general agreement on what counts as a non-canonical compound. Still, research on non-canonical compounds has been dominated by different approaches to classification. Taking this into account, in next section, we provide certain general definitions of non-canonical compounds in the pertinent linguistic literature.

¹ I am grateful to Professor Vesna Polovina for bringing non-canonical compounds to my attention.

2. THE GENERAL DEFINITIONS OF NON-CANONICAL COMPOUNDS IN LINGUISTIC LITERATURE

Many linguists agree that compounding is a highly productive process, and some of them even claim that "compounding is a primary source of *new* vocabulary" [my emphasis added] (Lardiere, 2006: 77). More recently, the process of compounding has been delimited by some authors as "[...] the concatenation of two (or more) lexemes to form a single new lexeme" and is "[...] a good example of a derivational process" because "compounding always results in the creation of a new lexeme" (Lardiere, 2006: 77). However, there are linguists who, for example, claim that compounding differs from derivation "in a way that is straightforward and traditional" (Anderson, 1989: 187). We have focussed on authors who do not impose limitations with regard to compound components number (cf. Chomsky and Halle, 1991; Copestake and Briscoe, 2005; Harley, 2009; Lardiere, 2006; Marsh, 1984; Master, 2003).

In the past few decades or so, non-canonical compounds have been classified in linguistic literature under different labels, such as: *complex noun phrases* (Marsh, 1984), *syntactic words* (Di Sciullo and Williams, 1987), *complex words* or *compounds* (Fromkin and Rodman, 1983; Julien, 2002), *complex noun sequences* (Marsh, 1984), *complex listemes* (Di Sciullo and Williams, 1987), *complex compounds* (Master, 2003), *noun sequences* (Vanderwende, 1994; Vanderwende, 1995), *complex nominals* (Levi, 1978), *long noun sequences* (Arens, Granacki and Parker, 1987), *multiword expressions* (Copestake and Briscoe, 2005; Harley, 2009), *multiple noun sequences* (Arens, Granacki and Parker, 1987), *nominal compound constructions* (Johnston and Busa, 1999), to name just a few. The main reason for this variety of terms used in linguistic literature might be the fact that even though typical compounds, or the most frequent compounds in general language are canonical compounds, the authors are, nevertheless, aware of the existence of non-canonical compounds and, therefore, look for some flexibility in terminology as well.

From what we have seen so far, we may assume that differences in classification seem to be merely a symptom of a more serious problem. In addition to this, different classification approaches do not illuminate, but rather obfuscate the phenomenon under consideration. That is why we opt for the term *non-canonical compound*. Although the term 'non-canonical compound' does not have the widest currency, it seems to us that it has the least restricted range of application. Therefore, 'non-canonical compound' seems to be a convenient cover term that enables us to include a broad variety of linguistic items under a single conceptual umbrella. In other words, by choosing the term 'non-canonical compound' instead of many other terms with narrower meaning we do not run the risk of evoking unintended connotations.

Strictly speaking, many authors simply do not define compounds in terms of the precise number of constituents, but rather define them either as "[...] words that contain more than one lexical morpheme" (Napoli 1996: 229), or as "[...] derived form[s] resulting from the combination of two or more lexemes" (Aronoff and Fudeman 2005: 236). Some linguists even claim that "[t]here is no theoretical limit to the lengths of compounds because the process of forming compounds can feed itself *ad infinitum* [...]" (Radford et al. 1999: 171-172). Also, Fromkin and Rodman (1983: 121) assert that "[n]ew words may be formed by stringing together other words to create compound words. There is almost no limit on the kinds of combinations that occur [...]."

Several questions can be posed in connection with non-canonical compounds then. The most basic one would be: Do non-canonical compounds share typical syntactic, semantic and phonological features that are generally ascribed to canonical compounds? In the following sections of our paper, we shall try to present certain similarities and differences between canonical and non-canonical compounds. In what follows, we summarise fixed order and recursiveness of compounds.

3. FIXED ORDER AND RECURSIVENESS OF COMPOUND CONSTITUENTS

Whilst observing syntactic features of compounds, many linguists take into consideration the following: the fixed order of constituents² within a compound and recursiveness.

It has been noted in the pertinent linguistic literature that the main difference between compounds and noun phrases is explained by fixed order of elements in a compound as opposed to the relative flexibility of constituents in a noun phrase. Another syntactic feature of compounds, which is mentioned in the literature on the subject, is recursiveness. Namely, according to Radford et al. (1999: 171) "it is possible to form compounds out of compounds." According to these authors, a canonical compound *finance committee* can be further expanded into non-canonical compounds such as: *finance committee secretary*, *finance committee secretary election*, *finance committee secretary election scandal*, etc. The similar example is provided by Donna Lardiere (2006: 77) who first mentions the canonical compound *toenail clipper*, and then goes on to expand it into the following units: *toenail clipper accident*, *toenail clipper accident insurance*, *toenail clipper accident insurance company*, *toenail clipper accident insurance company employee*, *toenail clipper accident insurance company employee benefits*, etc.

Some of the above examples of non-canonical compounds require further analysis. Moreover, we may further investigate whether *finance committee secretary election scandal* and *toenail clipper accident insurance company employee benefits* are compound words or syntactic phrases. For example, Donna Lardiere (2006: 77) states that "[e]veryone seems to agree that *toenail* is a word and even *toenail clipper* still seems pretty wordlike," and then, quite correctly, poses the following question: "Is there some definable point at which complex words lose their wordhood status and become syntactic phrase?" Even though this question seems to be a reasonable one, we may find further arguments in linguistic literature confirming the fixation of constituents within non-canonical compounds.

Fixed order of constituents and recursiveness of compounds can be illustrated by many examples from English specific discourses, for example, from the electrical engineering discourse³. In what follows, we shall illustrate non-canonical compounds with three and four constituents in such contexts.

² We assign very wide and liberal meaning to the term *constituent*, without any prescriptive intention, and without any theoretical commitment. Namely, in our paper the term 'constituent' refers to any item that linguistically functions as a unit.

³ I am basing my observation on the corpus that consists of spoken and written electrical engineering discourse, which was collected for purposes of Đurić (2012).

The same rule of fixed order of constituents is applied in English electrical engineering discourse for the three-constituent compound *diode bridge rectifier*, which cannot be reordered into **bridge rectifier diode*, or **rectifier diode bridge*. Finally, an example of four-constituent compound is *high voltage direct current*, which cannot be reordered into **direct current high voltage*, or **direct high voltage current*.⁴

As regards recursiveness, we might say that, sometimes, structural ambiguity grows exponentially upon expansion of constituents. In addition to this, certain authors claim that structural ambiguity occurs immediately in cases of non-canonical compounds comprising three constituents, for in such cases it is already unclear whether the first modifier in a sequence modifies the whole compound or the first constituent of a compound (Giegerich, 2009). Despite that, specialised discourses, as is the case of electrical engineering discourse, abound in examples of recursively applied compounding. A typical example illustrating this would be the case of a canonical compound *bridge rectifier*, which is further expanded into following non-canonical compounds: *diode bridge rectifier*, *three-phase diode bridge rectifier*, *low-harmonic three-phase diode bridge rectifier*. As can be seen from the previous example, we started with the two-constituent compound (*bridge rectifier*) only to end up with seven-constituent compound (*low-harmonic three-phase diode bridge rectifier*).

Examples such as these can be easily found in specialised discourse of electrical engineering⁵. However, the syntactic links and semantic relations between such compounds growing in number of constituents are not always easily determined. For example, the three-constituent compound *diode bridge rectifier* refers to a sort of bridge rectifier in connection with diode. Having added two constituents (*three* and *phase*), this compound becomes the five-constituent compound *three-phase diode bridge rectifier*: this is a type of bridge rectifier in connection with diode and simultaneously it is a three phase one. Then, if we fill an additional slot with the items *low* and *harmonic*, then we generate the non-canonical compound *low-harmonic three-phase diode bridge rectifier* that comprises seven constituents.

Paraphrasing can be here a practical and useful strategy that can assist in the retrieval of the implicit meaning, implicit prepositions, implicit relative pronouns or any other syntactic device which can explain the meaning and derivation of the given compound. Besides, paraphrasing is often used in linguistic literature as analytical device (Johnston and Busa, 1999). However, further investigation of the syntactic links and semantic relations remains yet to be done⁶.

The analysed examples raise an interesting issue of whether non-canonical compounds may be handled in semantic terms. In next section, we shall confine our discussion to semantic compositionality of non-canonical compounds in the discourse of electrical engineering.

⁴ See also Di Sciullo (2005) for an explanation of reordering illustrated by pertinent examples.

⁵ In no way does the imminent analysis rule out the possibility of these two syntactic features being used as non-canonical compound indicators in other discourse types. Nor does it aim at any prescriptive correction.

⁶ Unfortunately, the prosodic and acoustic criteria, the topics no less interesting, could not be taken up in this paper. Acoustic and prosodic aspects of non-canonical compounds are treated in Polovina and Đurić (2012).

4. SEMANTIC COMPOSITIONALITY OF NON-CANONICAL COMPOUNDS

In this section we first observe semantic compositionality of non-canonical compounds in General English, and then in electrical engineering discourse.

Although a compound is supposed to be one lexical item, created to have a specific, unique but generally compositional meaning, this is often not the case. Certain authors claim that non-canonical compounds are compositional to a certain extent, but the meaning is not predictable only on the basis of the meanings of individual compound components (Fabb, 1998). Often, the semantic interpretation of non-canonical compounds is difficult because the interpretation of meaning of these linguistic items might be influenced by a set of context-dependent and pragmatic factors (Đurić, 2012; Lapata, 2002; Marsh, 1984; Selkirk, 1982).

According to Leech (1977: 221), we must bear in mind the semantic bridge that can also be vague, and often extraordinarily indirect, particularly with compounds formed of three or more constituents, i.e. non-canonical compounds. We apply Leech's generalisation whereby X and Y may refer to several components. Leech states that for many compounds X-Y seems to be the most general rule, i.e. X which is in connection with Y, and this rule is sufficient enough to account for all plausible interpretations, even in cases of non-canonical compounds.

This broad delimitation is supposed to account for the meaning of the following compounds from general English, such as: *hunger strike* and *gunboat diplomacy* (Leech, 1977). On the one hand, the lexeme *hunger strike* is not so transparent, and it means "a refusal to eat for a long time, usually by a prisoner protesting against something". On the other hand, *gunboat diplomacy* means "the threat by one country to use military power against another in order to make it agree to something."

The same idea of compositionality may be applied to non-canonical compounds, most of which have transparent, compositional meaning, and those that are non-transparent, are non-compositional. In order to illustrate this distinction, let us see the following example. The three-constituent compound *electric power distribution* is transparent and fully compositional, for we can reconstruct its meaning. It is "a type of distribution of electric power." Similarly, the compound *scheduled maintenance time* is compositional, because its meaning is reconstructed as "time for maintenance which has been previously scheduled."

The neat examples so far cited are relatively straightforward cases of compositional non-canonical compounds. However, there are a whole lot of non-canonical compounds that do not easily lend themselves to such treatment, and are far from being semantically transparent. In other words, they are not compositional. This can be illustrated by the lexeme *master slave system*, which is not compositional, unless someone knows perfectly the context from which it has been taken. Namely, this semantic unit does not have anything to do with masters and slaves in the form of human beings, but rather refers to the specific technology which has unidirectional control over one or more other device. By the same token, the lexeme *Cheshire cat store* is not "a store that belongs to a Cheshire cat" but rather a sort of regenerative memory or dynamic memory.

A sketchy answer we are suggesting at this point might lie in the assumption that compositionality might not be the definite criterion for establishing the non-canonical compound status.⁷

⁷ This observation, however, merits further investigation.

Bearing in mind the fact that semantic compositionality and syntactic combinatorics cannot sufficiently account for non-canonical compounds phenomena, we introduce distributional features of non-canonical compounds as proposed by Aronoff and Fudeman, which, as we shall show, seem to cover our efforts to delimit non-canonical compounds.

5. DISTRIBUTIONAL EVIDENCE FOR NON-CANONICAL COMPOUNDS

The most consistent linguistic evidence that non-canonical compounds can be treated as such we derive from the work of Mark Aronoff and Kirsten Fudeman (2005: 106-107). This evidence is based on syntactic behaviour of these compounds and conforms with the ideas of fixed order of constituents within a compound on the one hand and, on the other hand, does not oppose different semantic criteria. The lexical integrity, which we have interpreted according to Aronoff and Fudeman (2005), is shown by the following three rules:

1. If the distribution of a non-canonical compound matches the distribution of any other noun, then this lexical item is a compound. The test is the following: If we can insert the given non-canonical compound into phrases like [a good + non-canonical compound], which is analogical to [a good + N], or if we can insert it into a phrase whose structure is [non-canonical compound + for hire] (analogical with [N + for hire]), then the non-canonical compound in question is a compound *par excellence*.
2. If a non-canonical compound behaves as a single unit for the purposes of *wh*-movement, then it has its own lexical integrity. Hence, the given non-canonical compound is a compound.
3. If the given non-canonical compound possesses the lexical integrity, then a modifier cannot be placed in order to modify such a unit.

We shall apply these criteria to some examples from our corpus. First, let us see the semantic unit *optically pumped solid-state laser*. Can we put that unit in a syntactic environment like [a good + N]? I bought a good *optically pumped solid-state laser*. We see that it behaves perfectly "normally" in terms of the English syntax. Can we say [N + for hire]? *Optically pumped solid-state laser* for hire. We see that it is also syntactically well-formed.

The *wh*-movement for this lexical item also supports its status as a lexical item:

Speaker A: Which solid-state laser did you see?

Speaker B: *The optically pumped one.

Speaker A: Which laser did you see?

Speaker B: *The optically pumped solid-state one.

Finally, we can apply the third test. Can we place a modifier *very* in this example *a *very* optically pumped solid-state laser? The answer is negative.

Having passed satisfactorily and successfully all three tests, we can assert that the lexical item *optically pumped solid-state laser* is a non-canonical compound construction *par excellence*. The line of reasoning proposed by Aronoff and Fudeman (2005) might clarify the non-canonical compound status. In particular, we should like to adopt those findings, which have been attested in our data, as non-canonical compound status criteria.

Furthermore, it seems to us that distributional evidence seems to be more promising for establishing the non-canonical compound status.

6. CONCLUDING REMARKS

Non-canonical compounds are very important part of English in electrical engineering. Broadly speaking, The idea of condensing the information, which would normally be expressed in a clause or even in a sentence, has found a fertile ground in electrical engineering terminology and language, simultaneously supporting the need for language economy which exists in this discourse type. It goes without saying that the continuous processes of research and improvement, which characterise technical sciences, cause the creation of new non-canonical compounds, sometimes almost on a monthly basis. It should also be stressed that their correct interpretation and understanding sometimes become very ambiguous and difficult.

In this paper, it has been claimed that the interpretation of English non-canonical compounds in electrical engineering discourse is not something that can be done easily, but rather, it is a procedure including certain theoretical assumptions and practical testing.

Some of the features of compounding as the process of creation of new words and compounds as the results of that process conform more easily to some well-established features of typical canonical compounds, especially when it concerns syntactic and distributional criteria. Semantic compositionality can also be invoked in many cases, but as the number of constituents grows, it becomes more difficult to decompose the meaning, as illustrated in our paper with the examples taken from the specialised discourse of electrical engineering.

This paper aims at promoting an interest in the study of non-canonical compounds, as there is ample evidence from corpus-based data that, sometimes, these items do not constitute a well-defined field within electrical engineering discourse and, more generally, across specialised discourses. Non-canonical compounds obviously need further investigation since most linguists agree that these items present the result of a productive process in language such as English, but as the number of constituents grows, it seems that most of the generally agreed theoretical assumptions on their status as compounds becomes less easily confirmed than in the cases when canonical compounds are examined.

REFERENCES

- Anderson, S. R. "Morphological Theory." In *Linguistics: The Cambridge Survey I - Linguistic Theory: Foundations*, edited by F. J. Newmeyer, 146–191. Cambridge: Cambridge University Press, 1989.
- Arens, Y., J. J. Granacki and A. C. Parker. "Phrasal Analysis of Long Noun Sequences." In *25th Annual Meeting of the Association for Computational Linguistics*, edited by D. E. Walker, 59–64. Stanford: Stanford University Press.
- Aronoff, M., and K. Fudeman. *What is Morphology?*. Malden, MA: Blackwell Publishing Ltd., 2005.
- Chomsky, N., and M. Halle. *The Sound Pattern of English*. The MIT Paperback Edition. Cambridge, Massachusetts: The MIT Press, 1991.
- Copestake, A., and E. J. Briscoe. "Noun Compounds Revisited." In *Charting a New Course: Natural Language Processing and Information Retrieval. Essays in Honour of Karen Spärck Jones*, 129–154. New York: Springer and Kluwer Academic Publishers, 2005.

- Di Sciullo, A. M. "Decomposing Compounds." *SKASE Journal of Theoretical Linguistics* 2(3) (2005): 14–33.
- Di Sciullo, A. M., and E. Williams. *On the Definition of Word*. Cambridge, Massachusetts: The MIT Press, 1987.
- Đurić, M. D. Složenice u diskursu elektrotehnike u engleskom i francuskom jeziku. Neobjavljena doktorska disertacija, odbranjena 12.12.2012. godine. Beograd: Filološki fakultet, 2012. [*Compounds in the Discourse of Electrical Engineering in English and French*. Unpublished doctoral dissertation defended on 12.12.2012. Belgrade: Faculty of Philology, University of Belgrade]
- Fabb, N. "Compounding." In *The Handbook of Morphology*, edited by A. Spencer and A. M. Zwicky, 66-83. Oxford: Blackwell, 1998.
- Fromkin, V., and R. Rodman. *An Introduction to Language. Third Edition*. New York: Holt, Rinehart and Winston, 1983.
- Giegerich, H. J. "Compounding and Lexicalism." In *The Oxford Handbook of Compounding*, edited by R. Lieber and P. Stekauer, 178-200. Oxford: Oxford University Press, 2009.
- Harley, H. "Compounding in Distributed Morphology." In *The Oxford Handbook of Compounding*, edited by R. Lieber and P. Stekauer, 129–144. Oxford: Oxford University Press.
- Johnston, M., and F. Busa. "The Compositional Interpretation of Compounds." In *Breadth and Depth of Semantics Lexicons*, edited by E. Viegas, 167–187. Dordrecht: Kluwer Academic, 1999.
- Julien, M. *Syntactic Heads and Word Formation*. Oxford: Oxford University Press, 2002.
- Lapata, M. "The Disambiguation of Nominalizations." *Computational Linguistics* 28(3) (2002): 357–388.
- Lardiere, D. "Words and Their Parts." In *An Introduction to Language and Linguistics*, edited by R. W. Fasold and J. Connor-Linton, 55–96. Cambridge: Cambridge University Press, 2006.
- Leech, G. *Semantics*. Harmondsworth: Penguin, 1977.
- Levi, J. N. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press, 1978.
- Marsh, E. "Computational Analysis of Complex Noun Phrases in Navy Messages." In *Proceedings of the 10th International Conference on Computational Linguistics*, edited by Y. Wilks, 505-508. Stanford: Stanford University Press, 1984.
- Master, P. "Noun Compounds and Compressed Definitions." *Forum* 41(3) (2003): 2–9.
- Napoli, J. D. *Linguistics: An Introduction*. Oxford: Oxford University Press, 1996.
- Polovina, V., and M. D. Đurić. *The Relevance of Stress in Delimiting Compounds*. [in print]
- Radford, A., M. Atkinson, D. Britain, H. Clashen and A. Spencer. *Linguistics: An Introduction*. Cambridge: Cambridge University Press, 1999.
- Selkirk, E. O. *The Syntax of Words*. Cambridge, Massachusetts: The MIT Press, 1982.
- Vanderwende, L. "Algorithm for Automatic Interpretation of Noun Sequences." In *Proceedings of the 15th International Conference on Computational Linguistics, Volume 2*, edited by M. Nagao and Y. Wilks, 782–788. Kyoto: Kyoto University Press, 1994.
- Vanderwende, L. *The Analysis of Noun Sequences Using Semantic Information Extracted from On-Line Dictionaries*. Redmond: Microsoft Research and Microsoft Corporation, 1995.