Compound Words' Classification - A Cognitive Linguistic Based Study

Padmaja Kilambi

University College of Arts & Social Sciences, Osmania University, Hyderabad, India

Email address:
kilambipadmaja@gmail.com

Abstract: This study attempts to classify compound words on the basis of Cognitive Linguistics and compares their usage trends using Computational Linguistics. In order to study compound words, it is very important to study the structure of a sentence because compound word in essence, is a condensed form of a sentence. After the Chomskyan Revolution, the concept of Cognitive linguistics in the structure of a sentence came into limelight. He explains about d-structure (deep structure), which determines the logic or meaning and s-structure (surface structure) that is the phonetic part. Lees, working with Transformational Generative Grammar (TGG), treated compound words, not as separate units but as a kind of embedded sentences and hinted for possible presence of d-structure and s-structure in compound words, which this study tries to investigate. Then on the basis of the Idealized Cognitive Model proposed by Lakoff and Fauconnier, compound words have been classified into transparent, opaque and counterintuitive compound words. Using Google Books Corpus, this study also compares their usage trends. This is done using usage frequency, defined in this work, which is analogous to productivity for affixed words calculated by G. E. Booij. Each class of compound word formed on the basis of ICM is found to have different usage frequency and the possible reasons for this are discussed.

Keywords: Deep Structure and Surface Structure, Idealized Cognitive Model (ICM),
Transformational Generative Grammar (TGG), Counterintuitive Compound Words, Usage Frequency

1. Introduction

Interest in word formation has started from the time of Panini, who provided a detailed description of Sanskrit word formations. Many questions on this subject were asked by scholars from 18th and 19th century [6]. The fact that words define language and that they are the main components of vocabulary, is neglected in the study of language for many years [6]. Even in descriptive studies, word formation was not paid much attention. American structuralists like Leonard Bloomfield [8] did not give importance to it because they were interested in units smaller than words - the morphemes. So, word formation was not given importance in structuralism theory [1]. Transformational grammar on the other hand dealt with units larger than the words like phrases and sentences. Therefore, the subject of word formation is considered a sub ordinate topic as there is no single theory on it and there is no agreement or relevant data to construct such a theory [6].

A major change in linguistics was marked by Chomskyan revolution in which he mentions that "…an attempt to characterize precisely the system of rules we have mastered that enables us to understand new sentences and produce a new sentence on an appropriate occasion will quickly dispel any dogmatism on this matter [13]."

Word formation remained a topic neglected by linguists, because Noam Chomsky focused on sentences and not on word formations. As Esko Pennanen [45] rightly points out, "…it is an obvious gap in transformational grammar not to have made provision for treating word formation." Robert Lees [35], who worked with TGG, treated compound words as special kind of embedded sentence [1] and with the publication of Chomsky's paper 'Remarks on Nominalization, [14]' the study of word formation became important.

A study of new words over the last fifty years (1941-1991) which was done by John Algeo [2-3] found that the percentage of new words from word formations is:

Compounding	40%
Affixation	28%
Shifting	17%
Shortening	8%
Blending	5%
Borrowing	2%
Others	0.5%

The above findings reveal that compounding and affixation account for two-thirds of the new words.

Compounding has the highest affinity to be created because of their tendency to combine and get stringed with different words. A couple of morphemes in a compound word can express any thought, feeling, or an opinion that one's mind can think of. Thus the study of compound words involves the study of language to a large extent. Chomsky [12] in his work proposed that "…the syntactic component of a grammar must specify for each sentence, a deep structure that determines its semantic interpretation and a surface structure that determines its phonetic interpretation."

Chomsky thus explains about surface (s-) structure, which determines the phonetic part, and deep (d-) structure that determines the logic or meaning in a sentence and other formal structures. Lees, in his work 'The Grammar of English Nominalizations' treated words, particularly compound words, not as separate units but as a kind of embedded sentences and hinted for possible presence of d-structure and s-structure in compound words, which this study tries to investigate. Chomsky hopes that through this concept of an underlying d-structure, we will be able to explain our intuitive understanding of many aspects of language [11]. If this is true for sentences in naturally evolving languages, it might be applicable for compound words also because they are condensed form of a sentence. TGG deals with intuitive understanding of a sentence. To understand compound words in a cognitive framework, a TGG based study is done. Further, in this study, they are classified depending on the level of intuition involvement in understanding them. ICM proposed by cognitive linguists George Lakoff [32] and Gilles Fauconnier [17] is taken as the basis for this classification. The usage trends of the classes of compound words are then observed by using Google Books Ngram Viewer and the results are analyzed.

The paper is organized as follows. After defining compound words in Sec. 2, some of the traditional classifications are mentioned. A study of compound words based on Transformational Generative Grammar which reviews Chomsky's description is done in Sec. 3. The influence of Idealized Cognitive Model (ICM), on classification of compound words, which is the basis of this work, is discussed in Sec. 4. A broader classification of compound words based on ICM is done in Sec. 5. In Sec. 6, the methodology is explained and usage frequency is defined. Sec. 7 deals with observation of usage frequency of classes of sample compound words followed by analyzing their trends. This study concludes with a discussion in Sec. 8, where some interesting possibilities for further research are indicated.

This paper attempts to answer the following questions.
1. How TGG supports the study of compound words in cognitive linguistic point of view?
2. Why is the classification of compound words in the light of ICM more reliable?
3. How is the classification by ICM different from other classifications?
4. Why some compound words live, some 'unlive,' and some die and what is the role of social and cultural factors on their observed trends?

2. Compound Word Formations

Much of the word formations in English is regular and is a rule governed creativity, yet, due to the fluidness of the word or due to rule changed causal usage; these patterns intermix and form hybrids [46]. This feature is observed in creative and productive mechanism of compound word formation because the users, unaware of word formation rules creates and uses them in their own context.

Definition

One of the greatest challenges faced by the linguist is to propose a rule to form words and their combinations. Often, it is observed that some compound words seem to be straight forward in their meaning, but turn out to be problematic upon close inspection. So, many theories were put forth to define the morphological aspect of compound word. The common definition is that compound words are morphologically complex words in which two words combine and yield a new meaning. One of the earliest definitions of compounds was given by Richard Morris [41] "Two or more words are joined together to make a single term expressing a new notion." Reka Benczes [7] defined compounds as "a combination of two words to form a new word." The form of a compound may be different when it is found in different context; they may be written as separate words, as a single word, or as being connected with a hyphen. Jonathan Charteris-Black [10] explains such differentiation from the viewpoint that how recently a word has entered the language. Novel compounds are usually written as two separate words and established compounds are written as a single word or hyphenated words. Compound words are usually made of two elements - the head and the modifier. The element that determines the semantic category is the head. It gives the meaning to the word. The element which modifies the head is the modifier. It adds to the meaning of the head word. For example, in the word '*hand bag*' the meaning is given by the element 'bag' because this word is referring to a bag, not the hand. So it is the head. The other element 'hand' is describing the head, so it is the modifier.

Generally, compound words like appositional (have two contrary attributes that classify the compound. E.g., *maidservant*), copulative (compounds with two semantic heads E.g., *sleepwalk*), endocentric etc., have a distinct head. The position of the head may be on the right side as in *wheelchair*, *overreact*, *afterthought* or on the left side as in

haircut, eyesight, nationwide. In some compounds, neither of its components can be categorized distinctly as a head, and its meaning often cannot be guessed immediately from its constituent parts. Such compounds are called exocentric compounds. E.g., *white collar, airhead, paleface.*

Compound words can be seen to be used in all kinds of contexts, such as newspaper articles, daily communication, and so on. It is one way for English speakers to create new English words-by putting existing words together. Though compounding is the most productive process of word formations in English, it opens many questions and problems which have not been solved yet [54]. Compounds are interesting, and at the same time intriguing and controversial linguistic constructions in terms of their analysis. They do not take a clearly determined position within grammar, since they connect several important linguistic and non-linguistic areas like syntagmatic and paradigmatic relations, syntax and morphology, and linguistic and pragmatic knowledge [49]. Some linguists (Algeo) perceive word formation as a grammatical rather than a lexical phenomenon, due to the productive power it has over other word formations. When it comes to compounding in particular, Algeo positions them on the border between lexis and grammar, and says that it is "part vocabulary and part syntax" [2]. So, there are many theories that classified compounds. Earlier Bloomfield classified compound words grammatically and Otto Jespersen [26] and Adams [1] semantically. Hans Marchand [40] could classify only exocentric compound words. In Elisabeth O. Selkirk [52] classification, each class of compound word is mingled with other. Though Judith N. Levi's [36] classification is most far reaching, it has many loopholes. For example, it could not explain classification based on taxonomy and it couldn't show marginal differentiation between exocentric and endocentric compounds. Although recent studies [4] showed that semantic transparency ranges continuously from fully transparent to fully opaque, some compounds words formed in modern times were not given proper place. Beatrice Warren [57] observed that the above classifications could not explicitly explain those compounds that lacked 'transparency' like *jailbird, pinkslip,* etc. To include such compounds, Benczes classified them on the basis of cognitive linguistics. This gave an insight to understand them by intuition. But there are some compounds that are not in the range of intuition. To address such words, compound words are classified on the basis of ICM.

By its nature, compounding provides insights into how the forms and meanings of individual morphemes map onto the form and the meaning of a complex word, and what impact this word formation has on cognitive demands of learning [51]. So, syntactic study of compound words based on TGG is important because it provides a cognitive approach to compound words. There are three reasons for this justification. Firstly, Lees, in his work 'The Grammar of English Nominalizations' treated compound words, not as separate units but as a kind of embedded sentences. This view was also supported by Bent Jacobsen [24]. Secondly,

the two constituents of compound words that exist now or had existed historically are mostly from open-class word sets in a language. Here, the stems or roots exist independently in the mental lexicon. So they can have numerous combinations. This characteristic of openness tends to create many combinations, like in a sentence as given by Chomsky in TGG. And lastly, Lees hinted for possible s- structure and d- structure in a compound word. Thus, compound words can be viewed syntactically and can be understood cognitively like a sentence.

3. Compound Words - A Transformational Generative Grammar (TGG) Based Study

TGG is, in the study of linguistics, a part of theory of generative grammar especially of the naturally evolved languages. Though he studied sentences here, TGG is applicable to compound words as they are miniature sentences. This can be explained by taking '*grief-stricken*' as an example. Its meaning can be interpreted as:

Someone is sad.

A tragedy struck someone's heart.

Someone is unhappy about what had happened.

TGG considers grammar to be a system of rules that generate exactly those combinations of words which form grammatical sentences of a given language. Chomsky believed that grammar has recursive rules allowing one to generate grammatically correct sentences over and over [59]. He explains it as "Transformational grammar is a device which generates only the well formed or grammatically correct sentences of a language since it is meant to create the rules and principles which are in the mind or brain of a native speaker."

This must also be true for compound words because they combine and generate exactly those words which form meaningful combinations [58]. If transformations of new sentences from the existing ones are possible, then new combinations of words to form new compound words from the existing ones are also possible. For example, the existing word *fear mongering* led to constructions like, *hate mongering, care mongering* and combinations like, *trustworthy, newsworthy, buzz worthy*, are other such productive analogies.

As Kumar Mitra [31] points out in phrase structure grammar (PSG), a sentence can have a sort of arrangement of its own component. This is called actual arrangement or its 'surface (s-) structure.' But Chomsky [13] says, "The grammar of a language is a system of rules that determines a certain pairing of sound and meaning. It consists of a syntactic component a semantic component, and a phonological component."

According to him sound and meaning join together in a language, in the same way d-structure and s-structure must join through some mental operations. So, there must be surface (phonology) and deep structure (meaning) to

compounds also. There might be some intuitive understanding – a mental process that goes on to understand the meaning of compound words by the head word. For example, consider

1. *Hair cut*
2. *Wheel chair*

When these two compound words are compared, in (1) we observe:

1. There are two morphemes as it should be in a compound word (s-structure).
2. The head is on the left hand side.
3. The meaning of the compound word is determined by the head word.
4. Therefore, the meaning is 'the hair that has to be cut' (d-structure).

Whereas in (2), we observe:

1. There are two morphemes (s-structure).
2. The head is on the right hand side.
3. The head determines the meaning of the compound word.
4. Therefore, the meaning is 'a chair that has wheels' (d-structure).

Chomsky, [14] in this regard, points out, "that deep structures which are often quite abstract exist and play a central role in the grammatical processes that we use in producing and interpreting sentences."| That means to get full semantic interpretation of a sentence we have to search for the underlying arrangement which he calls 'deep (d-) structure'. Also, Chomsky hopes that through this concept of an underlying d-structure, it is easy to explain our intuitive understanding of many aspects of language. "It is on this property that our inferences about mental process are based [31]." Thus, intuition is involved to understand sentences and hence to compound words.

To classify compound words, depending on the level of intuition involvement, Idealized Cognitive Model (ICM) is useful.

4. ICM and Its Influence on the Classification of Compound Words

ICM is the name given by cognitive linguists to describe a process in which knowledge is represented in a semantic frame work and it is a conceptualization of experience which is not in agreement with reality. It helps in understanding from the most concrete physical objects to the most abstract scientific concepts like science and language. It has been proposed by scholars such as Lakoff and Fauconnier who made attempts to show how ICM governs human thinking. Lakoff [32] considered ICM as a basic part of human ability to conceptualize. ICMs are "rich conceptual structures that capture relevant aspects of reality on the basis of a number of structuring principles [29]" and "they are highly abstract frames that can explicate some kind of typicality effect in the categorization [16]". Zoltan Kovecses [30] has demonstrated that verbal expressions and idioms used in language can be traced back to a limited number of conceptual metaphors. As

Bipin Indurkhya [23] points out, "…the metaphor that is placed as one of the elements in compound word opens up possibilities of establishing the meaning of the word through predictive analogy. It is the creation or inference of further similarities based on the previously existing ones which are predicted cognitively and thus its meaning is elucidated."

Lakoff and Fauconnier explained that the way we think, act and experience is a matter of metaphor. "Metaphors are linguistic expressions and are possible precisely because these are metaphoric in a person's conceptual system [33]." Metaphor is not just something related to poetry. It is the way to use new and interesting words by conceptualizing the similarities between two objects or words. Thus, the involvement of intuition is seen in such substitutions. Human thought processes and conceptual systems are largely metaphorical, structured and defined.

The paradigm of ICM consists of:

1. Prediction: The semantic pole of any linguistic expression (ibid; 97).
2. Meaning: the cognitive processing (ibid; 147).
3. Imagery: the ability to construe a conceived situation in many different ways (ibid; 47).
4. Context: the cognitive representation of the interaction between cognitive categories [55].

Cognitive model is the sum of experienced and stored contexts for a certain field by an individual (ibid; 55). Thus, they are helpful in understanding compound words that give direct meaning and those that can be perceived by prediction, imagery and context. John Vervaeke and Christopher D. Green [56] explaining about ICMs opined that these are linked to such diverse entities as particular words, modifying phrases, mental processes such as vision, scenarios, live individual and proposition. ICM can also be understood by Charles J. Fillmore concept of frames [18]. This concept explains that one cannot understand the meaning of a word without access to all the essential knowledge that relates to that word. For example, if we take the word *Wednesday,* it can be explained as the end of Tuesday, the beginning of Wednesday and the fourth day in a seven-day calendric cycle - a week. The frame here is the week system which has seven days and Wednesday is the fourth day.

David Rumelhart [48] has taken a propositional model to explain ICM, where schema is a network of nodes and links, every node in a schema would correspond to a conceptual category. The properties of the category depend on many factors like the role of that node in the given schema, its relation with the other nodes, relation among the schema and the overall interaction of that schema with other aspects of the conceptual system. There are three types of networks:

1. Node to node. E.g., *tablecloth*. It is about a cloth that is used for table not like a cloth in the shape of table or cloth that has a painting of tables on it. And a table is referred to as a wooden plank that has four or three wooden legs to stand on.
2. Node to schema. Eg. *clickbait*. Here, click is clicking the mouse and bait is to entice users to click the link. The click is connected to a node and bait has other schema as

in the word fish bait.

3. Schema to schema. E.g., in the word *White House*, the connection is between schema to schema as it is used to refer to 'the President of USA.'

Lakoff [32] distinguishes five types of ICMs:

1. Propositional ICMs: These include simple elements and the connection between them. E.g., sceneries and scripts.
2. Image - schematic ICMs: These describe different concepts based on the location of the objects, its shape and trajectory motion. E.g., latitude and longitude which represents location on the earth.
3. Metaphoric mapping: It deals with metaphors. E.g., comparing white to peace.
4. Metonymic ICMs: It deals with applying when one concept replaces the other. E.g., *New Delhi* referring to *the Government of India.*
5. Symbolic ICMs: These describe the knowledge about the language. E.g., grammar

ICMs are used as a way to represent knowledge structures and help to describe the semantics of linguistic expression. The cognitive approach to language focuses on how the mind conceptualizes our experiences of the world. As a result of its cognitive abilities, the use of psychological notions such as gestalt figure, ground, fame, attention, salience, intuition etc., are used to determine meaning. The classification of compound words which is done in this work took propositional, metaphoric, metonymic ICMs as reference.

5. Classification of Compound Words Based on ICM

Depending on the level of complexity of intuition involvement, this study categorized compound words as follows.

5.1. Transparent Compound Words

Transparent compound words include two elements whose meaning is transparent and one can understand its meaning even in a context free situation. According to Gary Libben et al. [38], English compound words that exhibit transparency derives its meaning from its elements. They have the highest semantic transparency. Based on ICM, the degree of complexity of intuition involvement is very less.
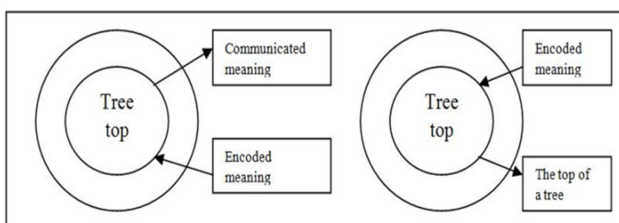

Figure 1. Effect of ICM on Transparent Compound Words.

There is no need of some special mental process to understand the meaning. For example, treetop is top of a tree as shown in figure 1. Other examples include earthshaking,

wanderlust, kindhearted, brainchild, blackbird etc.

5.2. Opaque Compound Words

The word opaque is taken from cognitive linguists Rene Dirven and Marjolijn Verspoor [15], who used the term in the study of metaphorical compounds and said that they cannot be analyzed easily. They contain two elements in which one or neither of them has a direct meaning. Their semantic transparency is lower than transparent compound words. Based on ICM, the degree of complexity of intuition involvement is higher. One has to use a higher degree of intuition to know their meaning. They are understood by their origin and with the similarities that they represent; if one word is understood, we get the response to another word without any special mental process. They are best understood in a context and the semantic meaning is outside the elements or they have to be inferred from the elements. For example, consider the word *greeneyed* as shown in figure 2. The head word is 'eyed.' The word intuitively refers to eyes that are green. But, the meaning of the word is also 'eyes that are filled with jealously.' To understand its meaning, one has to understand the metaphor that the word green represents. Thus, one can know what meaning can be inferred according to the context in which it is placed.
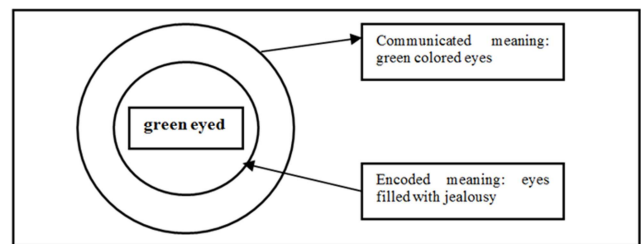

Figure 2. Effect of ICM on Opaque Compound Words.

Depending on the opacity of the two elements present, these are of the following types:

The head element of compound words is transparent, but the modifier is not transparent (OT). For example, '*gooseberry*' is a kind of berry but not directly related to the meaning of goose.

The head element in compound words is not transparent, but the modifier is transparent (TO). For example, '*jailbird*' is related to jail but not bird but the prisoner (which the word is meant for).

Neither of the elements is transparent (OO). For example, the word '*honeymoon,*' was started by the Babylonians who drank mead (honey beer) in the lunar month (moon). The word was 'honey month' which became 'honeymoon.' Thus, the meaning of both the constituents is opaque.

Based on the concepts of metaphor and metonymy, Benczes classified opaque compound words into.

5.2.1. Metaphor Based Compounds

In these compounds one of the elements is a metaphor. As Benczes [7] put it "…. the use of metaphor does not make the meaning of a construction more ambiguous than a "literal"

term, because if this were the case then they most probably had not been even coined in the first place with the help of metaphorical meaning extensions."

E.g., *jailbird*, *belly button*, *flame sandwich*, *greeneyed,* and similar others.

5.2.2. Metonymy Based Compounds

Though, it is an old-fashioned view of metonymy, where one word is used as metonymy for another word [5], these compounds are explained not only in words, but in thoughts and concepts, and across more domains than just language [39] and that are closely associated with the meaning of the word. They are best understood in a context. For example, in the word '*pink slip*' (a notice of dismissal from employment), one has to understand the metonymy of pink (Ford Motor (a company in America) issued a pink slip to denote that the employee has been terminated.) These are characterized by target domain where a profile determinant is absent and their study leads to what is called predictive analogy. The elements of these compounds are associated with the extension of the meaning of the domains. E.g*., cereal box, jackpot justice, chainsaw consultant, trophy child/wife, toy food,* etc.

5.2.3. Metaphor and Metonymy Based Compounds

Louis Goossens [20] created the word 'metaphtonymy' for these compounds. These compounds have constituents whose meaning cannot be determined with their metaphorical meaning. They are understood if placed in the context. Warren [57] gave the example of *jellybean* (a stupid person) for a compound word that has metaphor in metonymy. These compounds fall into large lexicon and are an instance of creativity. They are used extensively by poets and writers and are better understood when placed in a context. E.g., *bell bottom, muffintop.*

During the study, it has been found out that there are other types of opaque compound words called Proper noun based compound words which are entering into the active lexicon these days in which one of the elements is a proper noun.

5.2.4. Proper Noun Based Compound Words

In these set of words, either the head or the modifier is a proper noun. The characteristics of the modifier are determined by the characteristics of the proper noun which means that meaning of these words can be derived by knowing about the proper noun. They are usually the names of famous, notorious or of those people, places or things that caused change either in the world or in the lives of people. These last as long as the effect lasts. E.g., *Byronic hero*, Trump's *filter, Lutyens media, grammar Nazi, Frankenfood,* etc.

In recent times, however some of the compound words that are coined do not seem to fit in either class given above. Hence, a new classification has been given based on ICM to describe such words. These are called counterintuitive compound words and as the name suggests, they cannot be understood by intuition.

5.3. Counterintuitive Compound Words

These words have two elements of which neither of them has direct meaning. It is very difficult to guess the meaning of both the constituents even if placed in a context. E.g*., smell fungus, scapegrace* etc. If we take the example, *gaslighted,* the following analysis can be made:
1. It may mean that something is as light as gas.
2. One is feeling as light as gas i.e., relaxed.
3. It may mean that one's soul is burnt.
4. Gas is lighted.

But, the meaning given by OED is 'Manipulate (someone) by psychological means into doubting their own sanity.' The intuitive meaning is blocked and the user cannot guess the meaning by productive analogy. In another example, in the word *smell fungus*, one can observe how the real meaning is unpredictable as shown in figure 3.
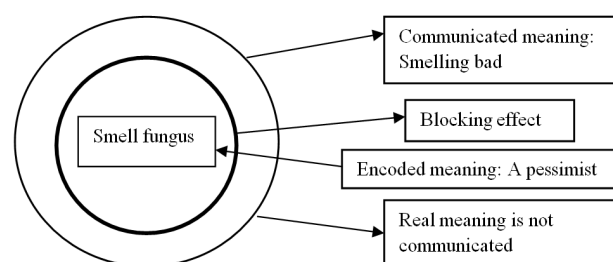

Figure 3. *Effect of ICM on Counterintuitive Compound Words.*

According to Philip B. Gough [21], context plays a role in the identification of words in text, and studies of context effects support the claim that words are recognized better in context than out of context. Paul Nation and James Coady's study [43] showed that the redundancy or richness of information provided in the context could enable a reader to guess an unknown word successfully without really knowing it with 'preconceived notions.' Shashi Tharoor [53], on improving vocabulary said that to understand the meaning and usage, one need not refer to a dictionary but rather look for the context in which it is used. That means if a word can be understood in relation to the context, its meaning can be predicted as in TCW. The little vagueness that OCWs carry may be related to what Plato said at the end of Cryatylus, in which he admitted that some social conventions were also involved, and that these were faults in the idea that phonemes had a separate meaning [50]. For example, words like *paper clipping, trickle ghosting* and *cause playing*, though convey a vague meaning, they can be understood to those people who are familiar with that particular jargon.

The main function of morphology is to expand the set of available words. When a word is formed and is used by more than one native speaker, and on various occasions, then the user will recognize that word which he had come across before and that established word will function as the lexical norm or lexical convention of that language. When a word is established, we say it is lexicalized. An important effect of lexicalization of compound words is that 'one of the elements of the compound may change in the meaning or disappear from usage [27].' The word that survives remains in the lexicon because of usage and any other word connected to it will be recognized by the user. This connectivity triggers the

formation of new compound words. These new words can be understood on the basis of the 'survived' word which is already present in the lexicon through cognition. Hence, ICM is important.

6. Methodology - Computational Linguistics

Finding the usage frequency of different types of words using Google Books Corpus.

A corpus is a collection of data concerning actual language use, these days mainly in electronic form. It is better than a dictionary because a dictionary is always lagging behind with respect to the use of productive morphological patterns, because it only registers the established words. A corpus (plural corpora) or text corpus is a large and structured set of texts (nowadays usually electronically stored and processed). They are used to do statistical analysis, hypothesis testing, checking occurrences or validating linguistic rules within a specific language territory. Corpus studies have been used before in order to test theories in personality traits [47], gender studies [42] and intonations [22]. Here, we use Google Books Corpus [25] in order to test the usage of different classes of compound words. This corpus has a collection of 155 billion words (n-grams). Due to limitations on the size of the n-gram database, only matches found in at least 40 books are indexed. To find the usage frequency in a particular year, only the books written in that particular year are taken into account. This corpus consists of n-grams and their usage frequency (which is defined below) over a period of time. The data is available for download, and can also be viewed through the interactive Google Books N-gram Viewer at http://books.google.com/ngrams.

In the fields of computational linguistics and probability, an n-gram is defined as the contiguous sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. An n-gram of size 1 is referred to as a 'unigram;' for example, *lotus* and size 2 is referred to as 'bigram' E.g., *black bird* and so on.

When the user enters an n-gram in the N-gram Viewer, it displays a graph showing how they have occurred in a corpus of books over the selected years until 2019. These n-grams can be any word or word formations. The X-axis shows the year and the Y-axis shows the percentage of n-grams present in the corpus that are identical with the input that is entered by the user. This percentage, in this work is termed as the usage frequency. The percentage is given by:

$$\frac{Number\ of\ given\ ngram\ repeated\ in\ the\ corpus}{Total\ number\ of\ ngrams\ present\ in\ the\ corpus}\%$$

To make the graph smooth, the frequency at which a given n-gram is repeated is averaged over its neighboring years. For example, the value at 1989 is given by the average of 1988, 1989 and 1990.

Geert Evert Booij [9] calculated the productivity in the case of affixations as:

$$Productivity = type\ frequency / token\ frequency.$$

The type frequency is the count of number of different word types of a morphological class. For example, all un-prefixes like unable, unclear, undo, etc, belong to similar type. Token frequency is the summed frequency of all the words of that type. Token frequency will be more than type frequency because the individual word might have been used many times. Productivity cannot be extended to compound words because affixes can string with other words, whereas many compound words form idiosyncratic combinations. So, there is no type frequency and so the usage frequency is used instead.

7. Usage Trends – Observation and Analysis

A sample of words from each class of compound words is taken at random in order to study their usage frequency.

7.1. Transparent Compound Words (TCW)

Aim: To know the usage frequency of transparent compound words.

Sample words taken: Brainchild, flower pot, treetop, eyeliner.

Observation: There was a rise in the frequency of their usage over the last century. This can be clearly seen in figure 4, which is plotted from 1900-2019. Though there are some fluctuations in the usage, one can see that there is no noticeable fall of usage of any of the words belonging to this category.

Explanation: This is because the elements of these words have a meaning that is understood intuitively in a context free situation. It enables the user to carry out interpretation with least possible mental machinery and least possible requirements regarding language learning. As Libben [38] points out, the understanding of TCWs involves two levels.
1. Constituency.
2. Componentiality

For example, if the word eyeliner is taken, both the elements that are constituents are used in their original meaning. It is also componential because the meaning can be understood with individual elements. Such words are user friendly hence the usage is more as shown in figure 4.

7.2. Opaque Compound Words (OCW)

Aim: To know the usage frequency of opaque compound words.

Sample words taken: *nanny state* – metonymy based OCW, *greeneyed* – metaphor based OCW, *Franken food* – proper noun headed OCW, *cliff hanger* – metaphor metonymy based OCW.

Observation: The figure 5 shows that these words were not used extensively till 70's. But there is an increase in the usage

of these words after 70's.

Explanation: The reason might be that these are ornate words mostly used by writers and poets. Typologically, one of the elements is transparent and the other has a meaning that is noticeable according to intensity of exposure, relevance, familiarity and sometimes education of the user. The observed trend of the sample words can be explained as follows:
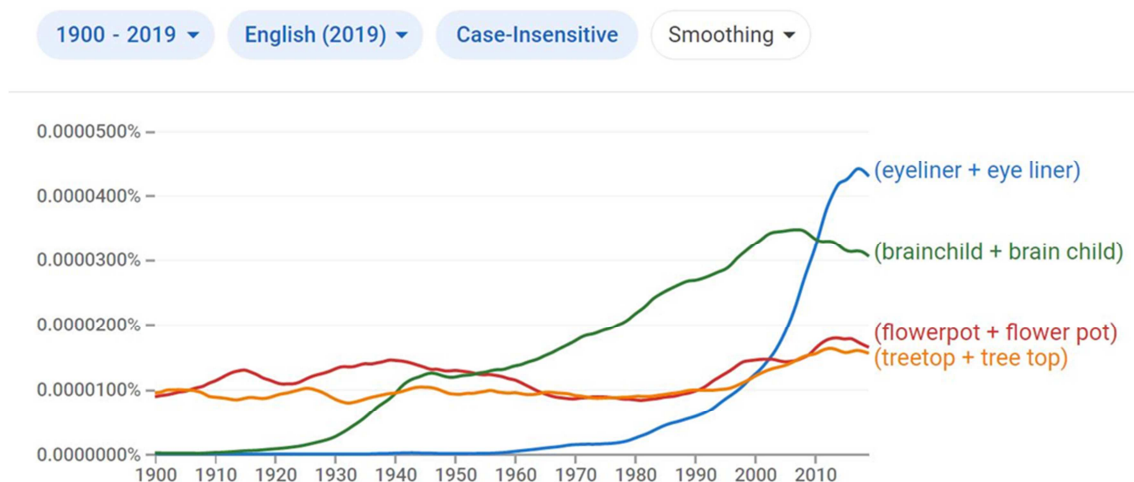

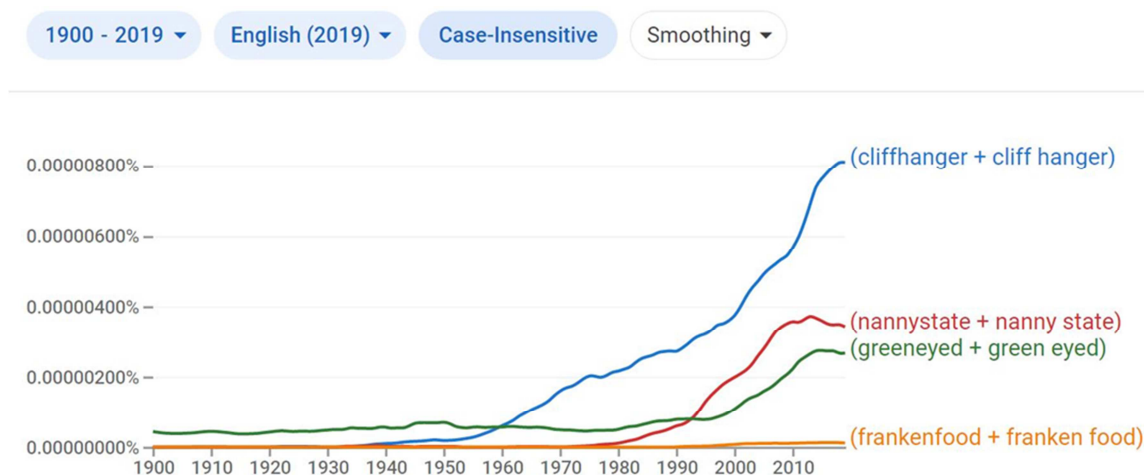*Figure 4. Graph showing usage of TCWs in different years.*

Figure 5. Graphs showing usage of OCW in different years.

1. The usage of cliff hanger became more due to increase in the ability of the user to co-relate with the situation of suspense or thrill in real life situations.
2. The word nanny state is popularized by the media to describe America and its relation with other countries. The widespread knowledge of world politics enabled the user to understand this word better.
3. The word greeneyed, though used by Shakespeare to describe jealousy, is also used in the context to describe someone with green eyes. The familiarity of green to jealous is decreasing because green is equated to the concepts like environment, prosperity etc. So people might have become comfortable in using the word jealous instead of greeneyed.
4. Franken food is synonymous with genetically modified food. With the effect of globalization and extensive rise in scientific temperament people got an exposure to the word Frankenstein. So a small rise in its usage is seen in the 21st

century.

7.3. Counterintuitive Compound Words (CICW)

Aim: To know the usage frequency of counterintuitive compound words.

Sample words taken: *Rain check, smells fungus, gas lighted*, and *scapegrace.*

Observation: Though figure 6 shows similar patterns as OCW, one must observe that there seems to be a fall in the usage frequency of these words and then a sudden rise in the 21st century.

Explanation: Both the elements together carry a meaning that is different to what the user predicts. The meaning they reflect is outside the deductive rationality. The elements are enclosed by a shell caused by blocking effect. Blocking effect is defined as "a phenomenon in which a previously learned thought process prevents or delays the learning and

conditioning of new behavior. It may also be a process wherein the flow of thought is obstructed or interrupted [44]." So, the user cannot guess the meaning because his intuition is blocked. The user cannot correlate, familiarize or contextualize the meaning because what is predicted is not what the word means. For example, the word smell fungus

has less usage frequency as shown in the graph because the meaning is counterintuitive (a pessimist).

According to the data collected, words like rain check is used in its literal meaning more than the meaning that it is really meant for. Here are a few instances where the word rain check was used in its literal sense:
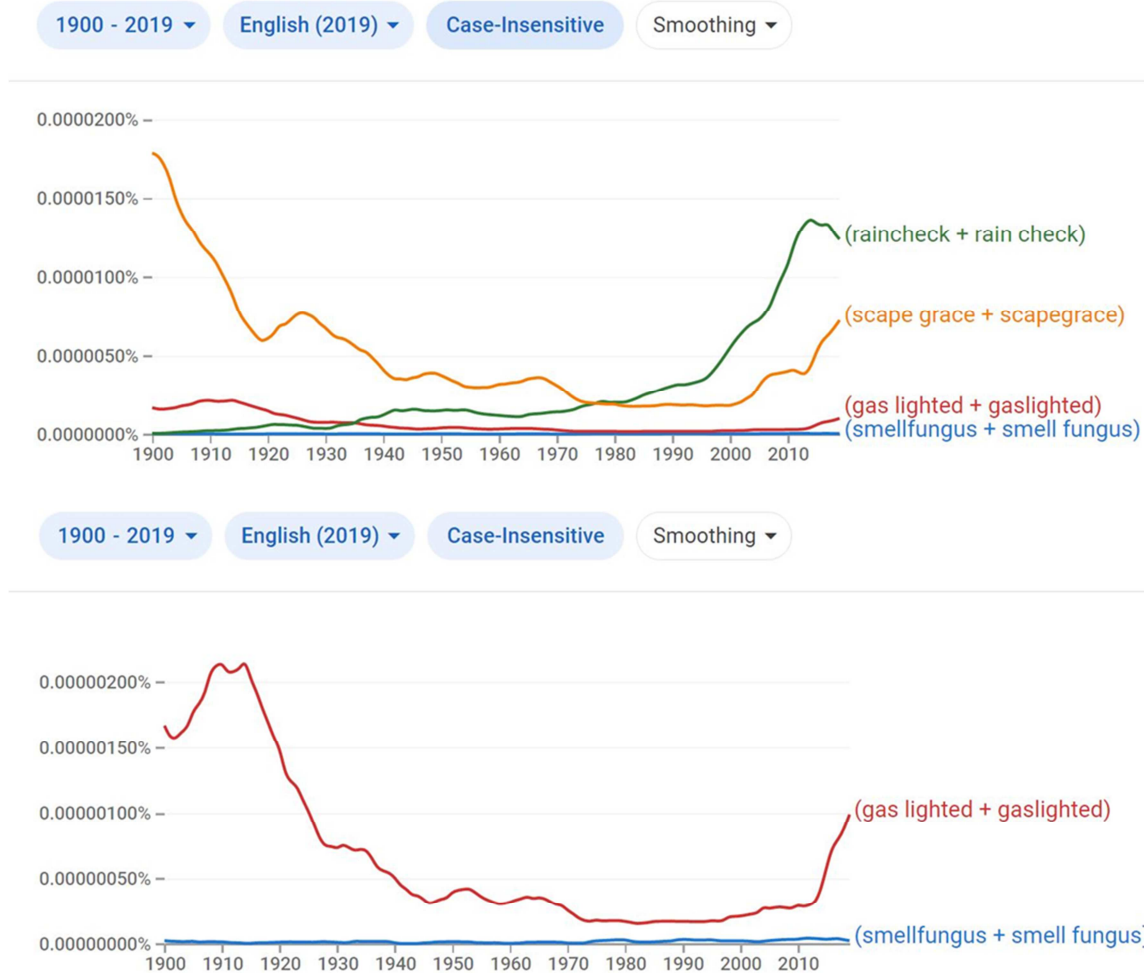

Figure 6. Graphs showing usage of CICW in different years.

1. "I'll take a rain check," she told him. A rain check she'd cash when hell froze over. (Heaven Only Knows, 1999) (OED: refuse an offer politely, with the implication that one may take it up at a later date).
2. When the rain stops and you don't want to carry the umbrella around drop in at the nearest stand — drug-store, cigar-store, hotel, theatre-lobby, and exchange your umbrella for a rain check again (Princeton Alumni Weekly, 1943) (literal sense: checking about rain).
3. Last summer on rainy days' people who purchased fifty cents worth of our goods at one time were given a rain check similar to the one reproduced below. (The New Idea: A Journal of Commercial Pharmacy, 1920) (OED: A ticket given for later use when a sporting fixture or other outdoor event is interrupted or postponed by rain).

Now, consider the word scapegrace (a mischievous boy). It was first used in the Arabian nights. As it has a

counterintuitive meaning, the usage frequency of the word faded because the user couldn't catch its actual meaning, but, since the 2000s, the usage frequency is increasing. One of the possible reasons can be that modern writers who write about fantasy and historical themes tend to use this word more often as observed in the corpus.

7.4. Comparing TCW, OCW and CICW

Aim: To compare the usage frequencies of three sets of TCWs, OCWs, and CICWs.

Sample words: *flowerpot* (TCW), *pink slip* (OCW), *Gas lighted* (CICW); *brainchild* (TCW), *cliffhanger* (OCW), *smell fungus* (CICW); *eyeliner* (TCW), *greeneyed* (OCW), *raincheck* (CICW).

Procedure: Three graphs, each containing the usage of each class of compound words are plotted as a function of time using the data available on the Google Books Corpus.

Observation: As seen in figure 7, the usage frequency of TCW is found out to be more than the usage of OCW, which in turn is found to be more than that of CICW. However, in the third graph, rain check, which is a CICW, is found out to have a higher usage frequency than green eyed, which is OCW.

Explanation: TCWs are user friendly, so their usage is generally more than the other two. The usage of OCWs is seen to increase with time as authors started using literary language and ornate words to express, describe and compare people or situations in a poetic way. As long as literature is appreciated by the readers, these words never die. CICWs have completely different meaning from the elements contained. The blocking effect makes the meaning of the word completely different from the meaning of the individual elements. The meaning of the word can neither be anticipated nor predicted. So the user tries to avoid them.
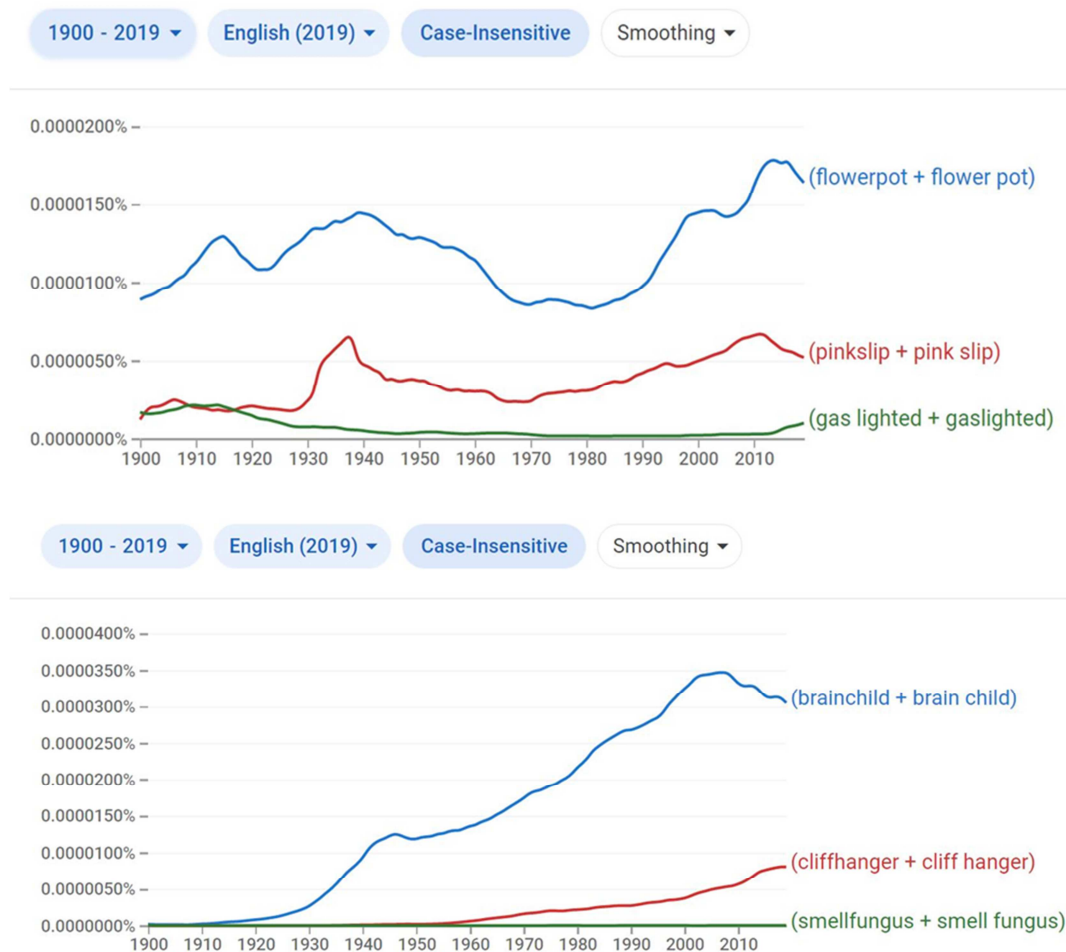
In figure 7, however, we have seen that the usage frequency of rain check, which is a CICW, is more than that of greeneyed, which is an OCW. The reason for this is that the user started using rain check in its literary meaning (to check the rain forecast). This is an evidence to say that there are some counterintuitive words which have an increase in their usage frequency because they are used in the context other than what they are made for. In such cases the word survives but the meaning dies.

The purpose of this study is to judge the utility of the word before creating it. English, being the 'lingua franca,' the user has a liberty to create those words that would satisfy his needs in the community he is living in. But language is not a crowd of words. It's a group of words. So, as a creator and user, one should take the responsibility of creating those words that are not only useful but also words that enhance the inexhaustible vocabulary of this ever demanding language.

The strength and validity of English lies in its readiness to welcome new words and expressions and to accept new meanings for old words, yet such words that do not support human cognition often depart as quickly as they arrive.

In his book 'The Tipping Point,' Macolm Gladwell [19] writes "the paradox of the epidemic:that in order to create one contagious movement, you often have to create many small movements first." The creation of new words results in such small movements which may cause a quake that would continuously evolve English. Out of the innumerable new words that are created, 40% are compound words. So, a study of compound words will influence the study of language to a large extent.
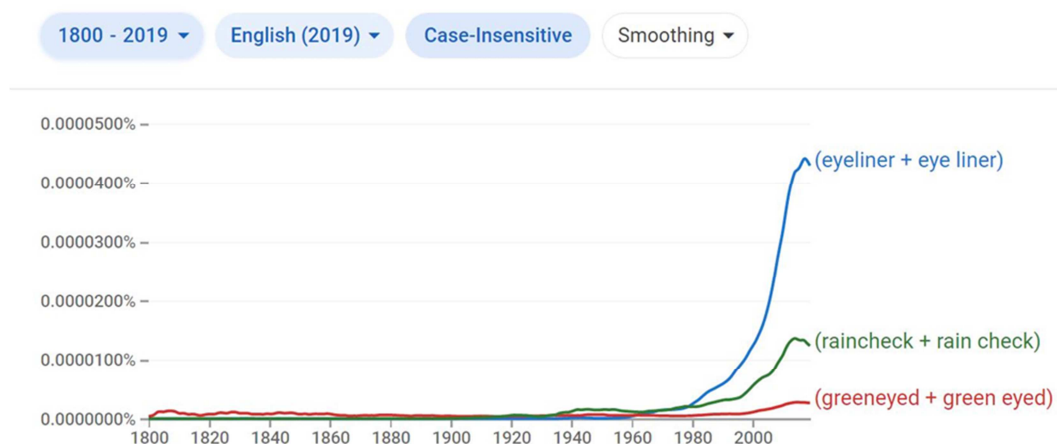
Figure 7. Graphs showing the usage of TCW, OCW and CICW in different years.

8. Conclusion

The study of language pattern under the light of TGG paved a new way in understanding compound words. Like in a sentence, compound words can also have s- and d-structures which are helpful in forming infinite number of combinations. ICM proposed by Lakoff and Fauconnier helped in understanding the degree of involvement of intuition, which led to the classification of compound words into transparent, opaque and counterintuitive.

To test their durability, a graph is plotted that showed different trends in usage frequency. Transparent compounds seem to be user friendly and are used extensively as the user feels comfortable because their meaning is transparent. These words have higher exposure and live in the active vocabulary forever. Opaque words on the other hand, are usually ornate and are used in literature. The elements have a concealed meaning and the reader should use higher degree of intuition to understand them. These words arrest the attention of the reader as they have high potential to express the deepest meaning. Proper noun based words like *Trump's filter, Franken food, Corona crisis* live as long as the cause for which they are coined lives. Words formed by productive analogy like *honey moon, yoga moon, child moon* provide instances where the user can become the creator. Counterintuitive compounds have the least usage frequency. They are not recorded in the active lexicon because their meaning is against intuition. User will not use these words because they are not stored in the lexical frame work. If the words are counterintuitive, then the intention and intuition fails to register them and in such cases the word becomes obsolete.

In the pre internet era, a word was accepted if it was coined by scholars. Now, due to language democracy [28] (freedom to use the language to suit one's need), a new word is created to describe an intention and it is understood by others through intuition. If the interpretation is not in accordance with intuition, then there cannot be any effective communication.

This work opens possibilities for future research. The following questions can be studied further.

1. Can we apply the studies of other cognitive linguists like Ronald Langacker [34] and Fauconnier in the study of compounding?
2. Can this paper open the study of other word formations like blending, affixation, clipping etc?
3. Is there any other parameter other than usage frequency which can be used to deduce whether the word lives, 'unlives' or dies?
4. Consider the possibility of using symbolic ICMs to study grammar.
5. What other linguistic forms can be studied on the basis of ICM?

References

[1] Adams, V. (2016). An Introduction to Modern English Word Formation. London: Routledge.

[2] Algeo, J. (2007). Vocabulary. In The Cambridge History of the English Language. Vol. 4, Edited by Romaine, Suzzane. Cambridge: Cambridge University Press.

[3] Algeo, J. (2010). The origins and development of the English language. Boston: Wadsworth Cengage learning.

[4] Auch, L., Gagne, C., & Spalding, T. (2020). Conceptualizing semantic transparency: A systematic analysis of semantic transparency measures in English Compound words. Methods in Psychology, 3, 100300.

[5] Bauer, L. (2017). Metonymy and the semantics of word-formation. Mediterranean Morphology Meetings, 11, 1-13. doi: https://doi.org/10.26220/mmm.2868.

[6] Bauer, L. (1983). English word-formation. Cambridge: Cambridge University.

[7] Benczes, R. (2006). Creative Compounding in English. Amsterdam: John Benjamins Publishing Company.

[8] Bloomfeld, L. (1933). Language. Chicago IL: The University of Chicago Press.

[9] Booij, G. E. (2005). The Grammar of Words: An Introduction to Linguistic Morphology. New York: Oxford University Press.

[10] Charteris-Black, J. (1998). Compound nouns and the acquisition of English Neologisms. (ERIC Document Reproduction Service No. ED 427 525). Retrieved from https://files.eric.ed.gov/fulltext/ED427525.pdf.

[11] Chomsky, N., Gallego, Á., & Ott, D. (2019). Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges. Catalan Journal of Linguistics, 0, 229-261. doi:https://doi.org/10.5565/rev/catjl.288.

[12] Chomsky, N. (2014). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

[13] Chomsky, N. (1969). Language and the Mind. (A. Bar-Adon, & W. Leopold, Eds.). New Jersey: Prentice Hall.

[14] Chomsky. N. (1968). Remarks on Nominalization. Indiana: Linguistics Club.

[15] Dirven, R., & Verspoor, M. (1998). Cognitive Exploration of Language and Linguistics. [Cognitive Linguistics in Practice]. Amsterdam: John Benjamin.

[16] Evans, V., & Green, M. (2006). Cognitive Linguistics: An Introduction. Edinburgh: Edinburgh University Press Ltd.

[17] Fauconnier, G. (1984). Mental Spaces: Aspects of meaning construction in natural language. Cambridge: Cambridge University Press.

[18] Fillmore, C. (1982). Frame Semantics. In Linguistic Society of Korea, ed., Linguistics in the Morning Calm. Seoul: Hanshin.

[19] Gladwell, M. (2000). The Tipping Point: How Little Things Can Make a Big Difference. Boston: Little, Brown and Company.

[20] Goossens, L. (1995). Metaphtonymy. The interaction of metaphor and metonymy in figurative expressions for linguistic action. In By Word of Mouth: Metaphor, metonymy and linguistic action in a cognitive linguistic perspective [Pragmatics and Beyond New Series 33]. Edited by Louis Goossens, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vanderbergen and Johan Vanparys. Amsterdam: John Benjamins, pp. 159–74.

[21] Gough, P. (1984). Word recognition. In P. D. Pearson (Ed.), Handbook of Reading Research. New York: Longman, 225-253.

[22] Hedberg, N., Sosa, J. M., & Görgülü, E. (2017). The meaning of intonation in yes-noquestions in American English: A corpus study, Corpus Linguistics and Linguistic Theory, 13 (2), 321-368. doi: https://doi.org/10.1515/cllt-2014-0020.

[23] Indurkhya, B. (1992). Metaphor and Cognition. Dordrecht, Boston: Kluwers.

[24] Jacobsen, B. (1977). Transformational-generative grammar: An introductory survey of its genesis and development. Amsterdam: North Holland Publishing Company.

[25] Jean-Baptiste, M., Yuan, K. S., Aiden, A. P., Adrian Veres, Gray, M., K., Brockman, W. The Google Books Team, Pickett, J., P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Martin, A., Nowak, and Aiden, E., L. (2010). Quantitative Analysis of Culture Using Millions of Digitized Books. Science.

[26] Jespersen, O. (1954). A Modern English Grammar on Historical Principles. London: Bradford and Dickens.

[27] Katamba, F. (Ed.). (2004). Morphology Critical Concepts in Linguistics (Vol. 2). Routledge.

[28] Kilambi, P. (2020). Deviation of English - A Subjective Study. The Literary herald, 5 (5), 19-26.

[29] Knop, S. D., & Rycker, T. D. (2008). Cognitive Approaches to Pedagogical Grammar. Berlin: Mouton De Gruyter.

[30] Kovecses, Z. (1986). Metaphors of anger, pride, and love. a lexical approach to the structure of concepts. Pragmatics and Beyond VII: 8, Amsterdam: John Benjamins.

[31] Kumar, M. (1999). How much deep are the 'deep structures'? From the Chomskian Perspective. Indian Philosophical Quaterly, 26 (3), 395-403.

[32] Lakoff, G. (1987). Women, Fire, and Dangerous Things: What categories reveal about the mind. Chicago, IL: The University of Chicago Press.

[33] Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. Chicago, IL: The University of Chicago Press.

[34] Langacker, R. W. (2008). Cognitive Grammar: A Basic Introduction. New York: Oxford University Press.

[35] Lees, R. B. (1968). The Grammar of English Nominalizations. 5th printing. The Hague: Mouton.

[36] Levi, J. N. (1978). The Syntax and Semantics of Complex Nominals. New York: NY: Academic Press.

[37] Libben, G. (1998). Semantic transparency in the processing of compounds Consequences for representation, processing and impairment. Brain and Language, 61 (1), 30-44.

[38] Libben, G., Gibson, M., Yoon, Y., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. Brain and Language, 84 (1), 50-64.

[39] Littlemore, J. (2015). Metonymy. Cambridge: Cambridge University Press.

[40] Marchand, H. (1969). The Categories and Types of Present-day English Word Formation (2nd ed). Munich: Beck, pp. 380-81.

[41] Morris, R. (1872). Historical Outlines of English Accidence. London: Macmillan and co.

[42] Motschenbacher, H., & Roivainen, E. (2020). Personality traits, adjectives and gender: integrating corpus linguistic and psychological approaches. Journal of Language and Discrimination, 4 (1), 16-50.

[43] Nation, I. & Coady, J. (1988). Vocabulary and reading. Edited by R. Carter & M. McCarthy. Vocabulary and Language Teaching. London: Longman, 97-110.

[44] Pam M. S. (2013, April 7). BLOCKING. Retrieved from https://psychologydictionary.org/blocking/.

[45] Pennanen, E. V. (1972). Current Views of Word-Formation, Neuphilologische Mitteilungen 73, 292-308.

[46] Plag, I. (2018). Word-formation in English. Cambridge: Cambridge University Press.

[47] Roivainen, E. (2020). Generational changes in personality: the evidence from corpus `linguistics. Psychological Reports, 123 (2), 325-340.

[48] Rumelhart, D. (1975). Notes on a Schema for Stories. In D. G. Bobrow and A. M. Collins (Eds.), Representation and Understanding: Studies in Cognitive Science (pp. 211-36). New York: Academic Press.

[49] Scalise, S. & Vogel, I. (2010). Cross-Disciplinary Issues in Compounding:Current Issues in Linguistic Theory 311. Amsterdam: John Benjamins Publishing Company.

[50] Sedley, D. N. (2003). Plato's Cratylus. Cambridge, UK: Cambridge University Press.

[51] Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44 (3), 421.

[52] Selkirk, E. O. (1982). The Syntax of Words (Linguistic Inquiry Monograph). Cambridge, MA: MIT Press.

[53] The Times of India (2019, November 11). A four letter word that defined Shashi Tharoor's famed vocabulary [Video]. YouTube. https://www.youtube.com/watch?v=nP5yqO5NX7I.

[54] Trips, C. (2009). Lexical Semantics and Diachronic Morphology. The development of –hood, - dom and –ship in the History of English. Tübingen: Niemeyer.

[55] Ungerer, F., & Schmid, H. J.(1996). An lntroduction to Cognitive Linguistics. London: Longman University Press.

[56] Vervaeke, J., & Green, C. D. (1997). Women, fire and dangerous theories. Metaphor and Symbol, 12, 59-80.

[57] Warren, B. (1992). Sense Developments. A contrastive study of the development of slang senses and novel standard senses in English [Stockholm Studies in English 80]. Stockholm: Almqvist and Wiksell International.

[58] Xiao, R. (2017). Combining transformative generative grammar and systemic functional grammar: Linguistic competence, syntax and second language acquisition. International Journal of English and Literature, 8 (4), 37-42.

[59] Yeh, R. (2011, November 13). Transformational Grammar. Noam Chomsky. http://noamchomsky2011.blogspot.com/2011/11/transformational-grammar.html.