Repetitions of Word Forms in Texts

Repetitions of Word Forms in Texts:
An Approach to Establishing Text Structure

By

Elena Tarasheva

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

TABLE OF CONTENTS

CHAPTER ONE

STUDIES OF REPETITION:
ESTABLISHING A UNIT OF ANALYSIS

A Review of studies of repetition

Like most things under the sun, interest in repetitions dates back to antiquity. Ancient Rhetoric starts off with a severe chastisement of repetition as a flaw in written style:

> Thus strings of unconnected words, and constant repetitions of words and phrases, are very properly condemned in written speeches: but not in spoken speeches—speakers use them freely, for they have a dramatic effect. In this repetition there must be variety of tone, paving the way, as it were, to dramatic effect. (Aristotle)

However, in *Poetics* Aristotle redeems the virtue of repetition as a figure of speech which purposefully strays from the every-day norm and gives essence to the elevated style–the most sophisticated among four language varieties. Following in his footsteps, Demetrius writes:

> The repetition of a word also conduces to elevation, as in the following passage of Herodotus: "There were huge serpents in the Caucasus, huge and many" (Vid. Herod. 1. 203). The reiteration of the word "huge" imparts certain impressiveness to the style. (Demetrius)

In modern day, a web search via the Google tool Scholar returns about 500 000 hits on the subject of repetition. Most results, however, are cognitive studies of the effect of repetition on memory and very few relate to style or poetics, as was the custom in antiquity. Indeed, on Scopus, one of the largest abstract and citation databases of research literature, "repetition" is keyword for more than 5 000 articles in the subject areas of neuroscience and psychology but only 1500–in the area of social sciences, which is the overarching rubric for rhetoric. Even there most of the articles dwell on

the link between repetitions and the functioning of the brain and not on how repetitions shape style.

Advice on using repetitions in writing and speech is particularly confusing. It varies from "Never ever" to "Avoid substituting one word for another only for the sake of variety". As prescription is rightfully out of fashion nowadays, we can expect some descriptive studies on the issue. However, none seems to exist. The closest that gets to such a study is the observation in Beaugrande and Dressler (1981) that in written texts repetitions would have been weeded out in the editing process, while the nature of the spoken language requires more frequent reminding of what is being talked about, and editing is impossible, so repetitions would be more frequent. No evidence is provided in favour of these observations and they are all too brief, so we cannot take them too seriously.

Discourse studies, as the branch in linguistics which research repetitions, shift the attention to repeating various units as a mechanism which makes a text "hang together". Two types of cohesive instruments are described– grammatical and lexical. The grammatical connectors accomplish some sort of substitution based on parameters proceeding from the respective language system–pronominal substitution, various deictics, the use of the definite article to point to a previous occurrence, overt comparisons, etc. However, they are not of interest to this study. Lexical links, which are in the focus here, are explained by Halliday and Hasan (1976) with language bonds such as synonymy, antonymy, or various semantic relations, like: the concept of "ordered sets" which connects ordinal numbers; the notion of "colour" which links adjectives like white, black, red etc.; relations of hyponymy, which unite mouth and face etc.

If one is to seek such relations in texts, however, serious obstacles occur. One impediment is polysemy. For instance, the sentences below both contain the verb "shoot", but in different meanings. Clearly, they are not connected into a cohesive text, despite the existence of this purportedly lexical link:

> *At least one protester has been **shot** dead and three wounded as thousands of Afghans demonstrated against plans by a radical US pastor to burn Korans on the anniversary of 9/11.*
> *Fifa chief is considering introducing the **shoot**-out after 90 minutes.*
> (*The Times* September 10th, 2010 front page)

A further complication to establishing lexical links is that there does not seem to exist a comprehensive list of lexical relations which perform text-formative functions. As in the definition above, an open enumeration is introduced, completed with etcetera. Researchers are thus left to their intuition to construe lexical relations obtaining between seemingly cohesive items in texts, while keeping in mind that some may be accidental. Clearly, such an approach is not conducive to rigorous analysis.

A different type of lexical mechanism for cohesion is proposed by Viehweger (1976). It is called "a nominative list" and is supposed to include all the lexis connected with a certain topic in a text, e.g. *snow*, *cold*, *winter*. The problem, of course, is whether we can include *scarf* and *sleigh* and what the respective text would look like:

1. *It became **cold** and **snow** started to fall.*
2. *Jack put on his **coat** and **scraf** and went to fetch his **sleigh**.*
3. *Models paraded wearing **scarves** and **sleighs**.*

Somehow 1 is more likely to be developed into a text by 2, rather than 3, although both contain the same parts of the prospective nominative list. Apart from the indiscriminate inclusiveness, such an approach to nominative lists poses the question: is it the case that the lexis is related inherently, or does the occurrence in a text make words cohere?

A more sophisticated approach is suggested by Morris and Hirst (1991:29), who employ the index of the thesaurus dictionary, which is based on a semantic classification. The index includes a number of general categories, such as "space", "matter", "intellect", subdivided into smaller categories, e.g. "matter" splits into "organic" and "inorganic". Each category gets a number, indexed also for the respective subcategories. Morris and Hirst propose 5 cases of linkage:

- Two words have a category common in their index entries, e.g. *residentialness* and *apartment* both have category 189: "presence";
- One word has a category in its index entry that contains a pointer to a category in the other word, e.g *car* has category 273 that contains a pointer to 276, which is a category of the word *driving*;
- A word is either a label in the other word's entry, or is in a category of the other word, e.g. *blind* has category 442, which contains the word *see*;

- Two words are in the same group and hence are semantically related, e.g. *blind–blindness*; *see*–vision;
- The two words have categories in their index entries that both point to a common category, e.g. *brutal* has index 851, *terrified* has index 860, and both have a pointer to 830.

The authors have obviously invested a lot of labour perusing the Thesaurus and this, indeed, appears a rigorous and principled classification. Even if other researchers are prepared to repeat the feat, however, they will not come up with the same results. Morris and Hirst used Roget's Thesaurus of 1977 and the edition of 1987 already has a different arrangement of the categories, therefore, different types of links will evolve from there.

The most straight-forward classification of cohesive lexical ties is proposed by Hoey (1991). It is based on two criteria–belonging to the lexical paradigm of a word, called "simple lexical repetition", or derived from the same root, dubbed "complex lexical repetition":

- Simple lexical repetition–the lexical item is repeated through forms of its grammatical paradigm, e.g. *bear–bears*. (p.52)

- Complex lexical repetition–the item and its repetition are derivatives but belong to different morphological classes, e.g. *drug* (n)– *drugging* (adj.)(p.55)

A third category presents what Hoey calls "paraphrases". They result from applying more than one type of lexical link simultaneously, for instance, *writer* and *writings* present a case of complex lexical repetition, on the one hand; *writer* and *author* are synonyms, on the other, and the two links lead to a third one—between author and writings. This is the so-called "triangle" (p.65) which bridges across different types of lexical relations. Such links may exist even when not all of their components are to be found in the text. However, despite falling within the category of lexical ties, some repetitions are called "chance" (p.56) because they do not perform linking functions. Thus a shadow is cast over the claim that lexical relations create links in texts.

All these studies are thorough and plausible, but any student of cohesive lexis will be rightfully confused by the complex procedures, the fuzzy principles and above all–by the direction of the search for cohesive lexis: is it the case that lexis has the power to link because of some inherent

(probably semantic) properties, or does the fact that it belongs to a topic introduced in the text make it so?

Another serious problem with studies of lexical ties is the inseparable relation between lexical links and co-reference. Having dubbed lexical cohesion "semantic", Halliday and Hasan (1976) classify it into expressions with identical, inclusive, exclusive or unrelated reference. While reference is defined as a grammatical tie, lexical links are supposed to be semantic in nature–therefore, substantiating a completely different type of link. However, most researchers end up looking for lexical relations among co-referring expressions in the text, as if no ties can exist without the process of co-reference. Now, if a relation of a specifically lexical character is sought, it should be divorced from the issue of reference. Otherwise textual connections of purely lexical essence can not be claimed to exist.

In a later work (1985), Halliday and Hasan adopt a classification based on the components in the naming process which will be described in detail later in this chapter. Ties which obtain between the extensions–the objects in reality named by the nouns–are known as "co-extension"; ties which are due to belonging to the same (morphological) class are known as "co-classification" (p.74); links which connect all the words used to refer to the same referent are known as "co-reference". In effect, co-extensions turn out to be synonyms, antonyms etc., co-classifications are substitutes, such as "so did he", "I am one" etc.; and co-reference is precisely the mechanism naming the same referent by nouns, pronouns and other expressions. Thus what the authors call "semantic principles" for lexical cohesion is entirely subsumed into the category of grammatical links, except for the co-extensions, which are again left as an unfinished list of presumably semantic relations. However, basing a classification of lexical links on the specifics of the naming process seems like a well-grounded approach and we shall adopt it for the study of lexical links. More about that will be revealed in the next part of this chapter.

A good point of departure for a study of repetitions is to ask the question: what is repeated? From the brief review of lexical cohesion above we saw that in some cases the reference to an object is repeated, in others – isolated semantic components, yet others reiterate merely the linguistic form. Needless to say, these are all repetitions of completely different types and should be treated in accordance with their nature. Furthermore, the unit of analysis differs in the literature on repetitions. While Demetrius's example quoted above discusses a single word—*huge*,

Aristotle points to the recurrence of syntactic structures which create parallels. Likewise, when writing about "repetition", various branches of modern Linguistics seem to refer to different units. Variation analysis, to begin with, looks at recurring structures which characterise the degree of formality (Schiffrin 1994:293). Speech act analysis, for its part, distinguishes between repeating the turns of each communicant or between the two communicants (Labov 1972:366). Genre analysis, for its part, concludes that advertisements repeat noun phrases to give prominence to the respective brand names, while product manuals repeat to ensure against mistaken identity (Cook 1994:32). These brief examples serve to show how different the units of repeating can be–from a conversational turn, through a noun phrase to a grammatical structure.

Therefore, a comprehensive study of repetitions should start with positing a unit of analysis which is:

- Undeniably lexical
- Reflects the process of naming
- Can be easily traced throughout the texts

To find such a unit of analysis, we explore the process of naming.

How lexical expressions refer to objects in reality

The process of naming involves linking three components–a lexical item (a word or phrase), an object in reality (a thing) and the concept of it. The sides in this process are traditionally represented as a triangle (Lyons 1968:404). However, for reasons which will become obvious later, we adopt a three-dimensional representation—the pyramid developed by Petofi (1985:89). It differs from the triangle in suggesting that the repetition might be of another form of the same word—the plural, or possessive, for instance. The semiotic pyramid, therefore, consists of a lexicon item, a form, a concept/ intension and a referent / extension (fig. 1). The lexicon item is a part of the vocabulary of a language; the form is a realisation of the item in speech bearing morphological and syntactic marking in accordance with the language system and the intentions of the language user. The concept or intension is a mental category which includes the every-day and specialised knowledge about the object. More complicated from the philosophical point of view is the apex "extension", or referent. Some notions, such as music, love, trust, etc. do not seem to single out any material object in reality, like the word "chair" would, for

example. Even more obvious is the case with verbs or adjectives, which rarely point to anything as concrete as the reference of nouns. Such terms will be considered non-referential here. They do not point to "things" in the world, but to concepts only.
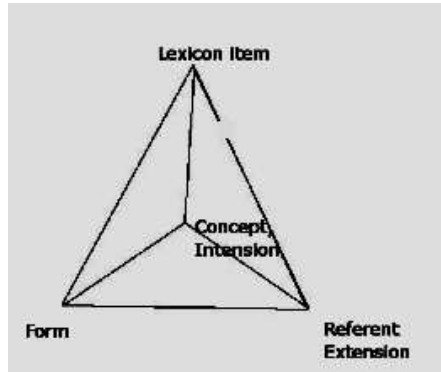
Figure 1. Semiotoic pyramid proposed by Petofi

The philosophy of language discusses issues of relating language to reality through the concept "description". It is defined as a language expression capable of evoking only one entity in the real world which satisfies some truth condition, e.g.

The King of France is bald

All expressions which name a king of France–if such a character existed-who is at the same time bald are truthful definite descriptions. (Russel 1905). While Russel's emphasis is more on the mathematical/ logical verification of the truth condition, Vendler (1967:36) sets out to establish which linguistic expressions perform the function of evoking one single referent. He calls such expressions "singular terms" and proposes that they evolve in a graded process. At the first stage an indefinite phrase is used relating the expression to a class and singling out one of its representatives. Secondly, co-reference to that representative is ensured by a second occurrence of the phrase with the definite article, thus confirming the identity of the referent. The process is schematically represented as follows:

There exists a N which ..

The primary introduction cannot include a generic phrase of the type:

A cat is an animal...

On the next stage of creating a singular term the definite article takes centre stage. The definite article is a function of a subordinate clause with a restrictive meaning. It shows that the scope of the noun thus restricted covers exhaustively all and every object within that scope. If the restriction covers a single representative, then the definite article is obligatory and it marks a singular term. In the other cases the term is generic and the definite article is not compulsory.

The fact that common nouns can name any of a class of objects is taken for granted in linguistic philosophy. It is the case that *man* can name all the individuals on the planet of male sex. Unlike a proper name, which normally refers to the individual it was given to at birth, a common noun can name a whole class, unless a definite description restricts the reference to one single object, e.g. *the man who came first in the competition*. Chesterman (1991:69-74) defines the referential set of a lexeme (U) as including two parts: objects or events which have been explicated in the concrete speech situation (r) and objects and events which are not part of the concrete situation, although they satisfy the reference of the noun (r^1). Each specific situation presents a different configuration of the set.

$$U= r+ r^1$$

Set r^1 is empty in cases of total reference (including all the members of the set) and U is equivalent to r. It is not empty in case of reference r^1 which is not total and U, therefore, is not equal to r.

This interpretation of reference can be extremely useful in cases such as these:

A boy climbed up the tree.
What is to be expected of a boy?

The intensions of the emphasised noun phrases are identical, the naming complexes coincide but the first phrase relates to **r**–a referent in a concrete situation. The second phrase, for its part, relates to **r^1**, because the referential set is not realised with the generic name. This is going to be a specific distinction in our treatment of repetitions–a phrase can be repeated with the referential set presented differently.

Another significant development in the philosophy of descriptions is the idea that some expressions are used to refer – to pick up a specific individual, while others focus on the description itself. To better understand the issue, take the famous example from Donnelan (1966):

*X wants to marry **a girl his parents disapprove of***.

When used referentially, a specific girl is named with the phrase "a girl his parents disapprove of" and it stands for the name given to this girl. Attributively, however, the expression does not single out one referent, but any girl X's parents disapprove of, i.e. specifies the characteristics of the desired candidate. Donnelan calls the two functions of noun phrases respectively "referential" and "attributive". In terms of our discussion here, we can say that the first type of use emphasises the extension, while the second–the intension. Therefore, the repetition does not evoke the same entity, but recalls a broader intensional content.

It is also customary in grammar to speak of "specified uses", where a representative of a class is envisaged, and "identifying uses", where a concrete person is referred to. Example 1 shows a specified use, while 2 – an identifying one:

> 1.   *I want **a rich husband**.*
> 2.   *That is **a man I met in the library yesterday**.*

A popular approach to unraveling meaning, following Frege, is to take as the fundamental unit of analysis the whole sentence. This is justified in view of the fact that the reference of a term is derivative from the entire proposition. Thus, phrases which can relate to objects in reality may happen to be used to refer to an empty set, that is – non-referentially. Allan (1986:144) describes four types of non-referential phrases in English:

- Phrases whose head falls within the scope of negation which presupposes or states that the denotation does not exist:

> ***God** does not exist.*
> *There are **no eggs** left.*

- Phrases whose denotation comes into existence with the respective predication or is expected to turn up later:

*You make **the toast** and I`ll make **the tea***
*You must write **a letter** to your parents.*

- Phrases where the denotation may or may not exist – the speaker does not know with certainty:

*Are there **any eggs** left?*
*Should there be some **eggs** left, we could have bacon and eggs.*

- Phrases whose denotation is an unspecified subset unknown to the speaker:

***Any dog** will get upset if you kick it.*

Later on in time Green (1989:40) summarises contexts which can create non-referential naming phrases. They include:

- conditional sentences
- modal verbs
- terms which create new worlds
- adverbial clauses

The final touch in describing links between words and reality comes with pragmatics, which deals with the intentions of the speakers. Pragmatics stresses the fact that the speaker has the right to select their reference within a freer range of options than the intension of a term gives. Nunberg (1978:6) introduces the notion of "pragmatic shifts" of the reference of a phrase. While *a ham sandwich* is a type of food, a waiter may use the linguistic expression to name a person who ordered such a sandwich:

***The ham sandwich** wants his bill.*

Likewise, when we say "I drank a whole bottle", we normally do not mean the glass, but the contents of that bottle, be it water, wine or cough medicine. The term "pragmatic shift of the intension" is, therefore, used for meaning relations, such as:" the content of ….", "the reason for…", "the publisher of …", "one type of ….", rather than the object itself. In effect, Nunberg claims that nothing limits the range and complexity of the functions performed in the course of the naming process, once the intentions of the speaker are taken into consideration.

Apart from pragmatic, intensional shifts can also be metaphoric. Several examples show that people can be referred to as "swines" or "lions", depending on which properties in the intension of the respective nouns are meant. In effect, the pragmatic component of naming takes us to Kripke's (1980) position that reference is causative, i.e. the speaker decides what characteristics to impart with their utterance. According to Reimer (2005), this approach is typical for linguistic descriptions of reference.

Finally, each object or event can be identified in one of two dimensions: the generic space, or the spacio-temporal (Thrane 1980:39). That is to say that the statement "*A tiger attacked the cheetah*" picks up two animals situated in the temporal dimension at a particular time and space. Contrarily, "*A tiger would attack his victim*" identifies one specimen of the genus in a non-specified situation, that is, generically. Another projection into the generic space would portray several animals as a group "*Tigers live in forests*", while yet another realisation in the same non-specified space would point to a typical representative of the set: "*The tiger is a carnivore*". Therefore, apart from using a phrase to refer, or to describe, the user can place their reference in a different dimension.

From the point of philosophy, involving various "worlds of existence" in the analysis of expressions belongs to the domain of modal logic, which differs significantly from the logic presented by the positivists quoted above–Donnelan, Frege etc. However, for linguistics reference in different worlds is necessary in view of the fact that literature often refers to parallel existences, other worlds etc. Furthermore, positivists would have to treat on a par the expressions "*a unicorn*" and "*Santa Clause*", whose referents do not exist in our world, with the referents of "*There are no eggs*" and "*Prince Charming might exist*", whose existence is denied or doubted. We claim that the uses of these–and other non-referential phrases differs and should be treated differently, which is why the concept of worlds will be allowed here, together with the concept of descriptions within various predications.

Conclusions

In conclusion, our review of the literature on repetitions revealed that the interest in repetitions as a feature of style is waning in recent years. We find this unreasonable because people do need to write and know whether to avoid repetitions, or employ them in specific ways. Further, the approaches to repetitions in linguistic disciplines are characterised by

confusion with co-referring terms, a lack of a unified unit of analysis and a fuzzy position on the issue whether lexical items link because they are semantically related, or whether a joint occurrence in a text causes them to be related in some way. Therefore, we set out on a search for a unit of analysing repetitions which would be undeniably lexical, easy to trace and proceeding from the way words refer.

In view of our criteria, the word form is perfect to perform the function of the unit of our analysis. We will trace whether the form, the intension and the extension change with repeated forms of a lexicon item. Because some notions are named by more than one word and the expressions may or may not be a fixed item in the nomenclature of a language, *in lieu* of "lexicon item", we will use the term "a naming complex".

The form can change within the grammatical paradigm of the repeated item–to show plural, singular or possessive; as the article is a separate word in English, we will also need to keep track of definiteness. When it comes to the extension, the speaker is allowed to pick up a specific representative of the set (an identified use), or to spell out its characteristics (a specified representative); to avoid naming a referent through a generic name, or to evoke an iterative representative; to classify through predicative uses, or to evoke a zero representative through negation. When repeated forms of words present the referent in one of these ways, we shall say that they have different referentiality, or reflect the referential set in a different manner. Additionally, the repeated items can pick up a different representative of the class and then we shall say that they have different referents, although they may be members of the same class. The research model also allows for establishing shifts in the intension–pragmatic or metaphorical.

In order to establish the role of repetitions as a text building mechanism, answers to the following five questions are sought:

> 1. How significant is the repetition of word forms in three different genres?
> 2. Is it the case that repeating word forms creates text structures?
> 3. What is the role of each side of the naming process: the referent, the thought and the form?
> 4. Does the nature of repetition change with each genre?

Three hypotheses are tested:

- The major role of repetitions is to recall the same object. If this is true, each recurring word form will evoke the same referent in reality.
- The main function of repetitions is to build conceptual links among the objects named. If that happens to be the case, the repetitions will evoke the same conceptual load while naming different objects, or remaining non-referential.
- The major role of repetitions is literary–to build figures of speech. In that case, the respective figures will be obvious when tracing the repetitions.

Corpora and method

The study was initially carried out using a software package called "The Linguist's Workbench" (Stambolieva 1996). Later, some of the analyses were repeated on the Wordsmith (Scot 1989), which provides richer statistical data and longer concordance lines.

The texts for the corpora are:

- Short stories written by established writers in English –61 125 running words in 9 stories.
- Research articles written by established researchers in English– 77 583 running words in 8 articles.
- Political speeches in English delivered by outstanding personalities–26 897 in 10 speeches.

Because the research needs complete texts, each one was run through the software independently of the others. To make the results from materials of different size compatible, a normalisation procedure was adapted from the one proposed by Biber (2000:263). The overall number of words is juxtaposed against the number of repetitions and a proportion per one million is calculated, using the formula:

number of repetitions/number of word forms X 1000

The result is called "index of repetitions" and is calculated for each text from each genre. The juxtaposition among the genres will reveal genre specifics; the juxtaposition among authors would show whether repetition is a matter of personal preference.

For comparative purposes, texts written by inexperienced writers are also included in the study. We proceed from the assumption that the parameters of the texts by established authors would reveal the essence of good writing; the features of the writing by inexperienced writers, for their part, would highlight differences, which would probably worsen the style as a whole.

CHAPTER TWO

REPETITIONS IN RESEARCH ARTICLES

In this chapter I establish the repetitions in an article written by an experienced researcher and try to elicit patterns: do the repetitions occur in phrases with the same type of reference; is the same concept evoked with each repetition; are there intensional shifts in the repeated items? Secondly, I check whether the established types of repetition chains can be found in other specimens of the same genre, the research article. Thirdly, I demonstrate common features of the repetition chains and answer the question how each type of repetition chain helps the author reveal his idea. Next, I show what type of repetition chains occur in the writing of inexperienced authors, students, and how the specifics of the repetitions lead to building a clear or deficient style. This serves as a demonstration how repetitions can be employed as objective indicators of good and bad writing. Finally, a way to summarise a research article is shown using my classification of repetition chains.

Part one: The repetitions in an article. Do they reveal a pattern?

When all the words from a text are arranged according to their frequency, we get the so-called "frequency list" (Baker et al. 2006). The job of counting and arranging is performed by special software products – in this case – the Word Smith Tools (Scot 1989). Below is the frequency list from Chomsky's article 'Linguistic Contributions to the Study of Mind' published on his website. The figure to the right presents the number of occurrences in this specific text. The first 90 positions are presented here as an illustration of the type of list acquired by these techniques, but the list is as long as the overall length of the article:

THE	881	AN	68
OF	811	BUT	66
TO	439	I	66
THAT	414	GRAMMAR	64
IN	336	THERE	64
A	314	CAN	61
AND	300	HAVE	58
IS	294	ONE	57
IT	173	FROM	56
FOR	149	OR	55
LANGUAGE	149	STUDY	53
THIS	136	STRUCTURE	52
AS	135	BEEN	51
BE	128	INNATE	48
ARE	107	NO	45
WE	90	THESE	45
HUMAN	80	HE	43
NOT	78	HIS	43
WITH	76	PROBLEM	43
HAS	75	THEORY	43
ON	73	AT	41
WHICH	72	KNOWLEDGE	39
BY	70	SUCH	39

As can be seen, the most frequent words are function words and words of broader meaning, as predicted by Zipf (1949). Because our interest here is in notional words–adjectives, nouns, verbs, adverbs and numerals, we first remove the function words–prepositions, articles, demonstratives, pro-forms, auxiliary verbs etc. Having removed those, the list becomes more manageable and these are the top entries:

LANGUAGE	149	LEARNING	33
HUMAN	80	LANGUAGES	32
GRAMMAR	64	POSSIBLE	32
STUDY	53	UNIVERSAL	32
STRUCTURE	52	GENERAL	30
INNATE	48	PROPERTIES	30
PROBLEM	43	SEEMS	30
THEORY	43	DATA	28
KNOWLEDGE	39	PRINCIPLES	28
MIND	36	EXAMPLE	27
SYSTEMS	35	ACQUISITION	26
SYSTEM	34	ANIMAL	24
FACT	33	CONDITIONS	24

FIRST	24		CASE	23
GENERATIVE	24		PSYCHOLOGY	23
LINGUISTIC	24		SPECIFIC	23
ORGANISATION	24		STRUCTURES	23
SENSE	24			

As has been noted by researchers (Scott and Tribble 2006, among numerous others), the most frequent items reveal the 'about-ness' of a text. In this case, the text is obviously about language, grammar, structures and the human mind. However, this is only an initial impression. If we look closer, the words can be sorted out into various groups.

Repetitions of *language*

The most frequent notional word *language* occurs more often than not in combination with the adjectives *human* and *natural*, and also in the phrase *the study of language*. The reference of the respective phrases tends to be generic, except for the latter phrase.

The notions evoked with the naming complexes include a number of interpretations of the concept:

1. *Anyone concerned with the study of human nature and human capacities must somehow come to grips with the fact that all normal humans acquire* **language**...
2. *... these studies simply bring out even more clearly the extent to which human* **language** *appears to be a unique phenomenon*
3. *In fact, it is difficult to see what links these stages at all (except for the metaphorical use of the term* "**language**")
4. *...the group which ought to have been able to evolve* **language** *in the true sense, and not the mammals*
5. **Language** *is purposive "in that there is nearly always in human speech a definite intention of getting something over to somebody else...*
6. *It seems clear that we must regard linguistic competence — knowledge of a* **language** *— as an abstract system underlying behaviour...*
7. *Such a grammar defines a* **language** *in the Humboldtian sense, namely as* "a recursively generated system ...

In the first example the author uses the meaning "power or faculty of speech", listed in position 4 in the Oxford English Dictionary (OED). In the second one the meaning is "the vocal sounds by which mammals and birds communicate", listed in the first position in OED. The third example

gives the citation form, which can pick up any of the meanings. Fourthly, language is presented as "the method of human communication, either spoken or written, consisting of the use of words in a structured and conventional way" (OED). In the fifth case the meaning "language (or a language) viewed as an abstract system, accepted universally within a speech community, in contrast to the actual linguistic behaviour or performance of individuals" is exploited. The sixth example envisages the most common meaning associated with the word: "the system of spoken or written communication used by a particular country, people, community, etc., typically consisting of words used within a regular grammatical and syntactic structure". Finally, a concept introduced by Humboldt is brought into the scene. The distinctions may appear minute, but they are clearly distinguishable – both according to the dictionary and in the respective uses in the text.

This pattern of occurrence indicates that the author deals with a number of concepts–other researcher's and his own–of the notion named with the word *language*. In terms of this study, we can say that the form is associated with a different part of the intension of the lexeme for each occurrence, while few actual referents are evoked. We can conclude that the repetition serves the purpose of discussing various understandings of the concept, as well as a range of aspects of the respective phenomenon. The shift in the conceptual content enriches the discussion and adds a range of viewpoints.

Another type of repetition presents the repeated word prefaced by various prepositional phrases:

> *knowledge of a **language**,*
> *the structure of a **language***
> *the study of **language***

In this way, the concept is picked up for reference in one of its specific aspects, rather than as a whole. I shall call this type of referring "restructuring", because the developed formal shape introduces a referent with a slightly changed identity. Therefore, the repetition does not stay steeped in the same unchanging intensional content but is enriched, thus pushing the presentation to further depth. An interesting feature of this type of reference is the co-existence of generic-ness and specific-ness. While *knowledge of a language* makes generic reference to knowledge of a specific language, *the structure of a language* calls a specific structure of

a language in its general sense. Thus we see a specific notion within a generic concept or a generic notion overarching a specific object.

Repetitions of *human*

A second type of notional word repeated frequently is *human*. Unlike *language*, this is an adjective and occurs in noun phrases with *language*, *intelligence*, *thought*, *psychology*, *freedom* and others. Its function is to relate the respective concepts to the sphere of human faculties and features. While the repetitions of *language* broaden the conceptual field with new aspects and modifications, the adjective *human* helps restrict the reference to a specific sphere.

Repetitions of *grammar*

The third highly frequent word is the noun *grammar*, which collocates with *innate*, *generative*, *transformational*, and *philosophical*. While the first noun we discussed was mainly included in generic phrases, this one occurs in quite a few specific phrases of a type which is illustrated with the following examples:

> *a set of data for this* **grammar** *to be confirmed*
> *that the* **grammar** *contains a phrase structure component*

The uses of *grammar* above can be characterised as substituting previously used phrases abbreviated in their second occurrence. This differs from other specific uses in the fact that the way the phrase projects its denotation includes a previous mention of an entire phrase. Such reference is useful in cases when qualifications are made of a previously mentioned object. I am going to speak of this referential type as a **substitute.**

Yet another type of phrases in which *grammar* is included is a number of generic phrases, illustrated with the following examples:

> *now available suggests that if universal* **grammar** *has serious defects as indeed it*
> *it is reasonable to suppose that a generative* **grammar** *is a system of many*
> *the* **grammar** *of a language must be discovered by the child*

The examples show three types of generic nouns: with the zero article, with the definite one and with the indefinite article, each-with the respective characteristics. All of them present the referential set differently: with no specific representative–in the first case, with a random

representative–in the second, and with a typical one–for the definite phrase.

The repetition is included as a substitute in the following case:

> *that enables it to construct such a **grammar** from the data of sense*

Identifying phrases contain repetitions of *grammar* in the following examples:

> *determining whether given data are compatible with a **grammar** of the given form*
> *a schema to which any particular **grammar** must conform*
> *what relation must hold between a potential **grammar** and a set of data*

Therefore, we can conclude that the repetitions of the noun do not change the intension of the noun, but present different referential types. As is known, generic nouns are instrumental in making generalisations, specific uses tend to exemplify issues or distinguish sub-types, the type we called "a substitute" bestows qualifications. That is why the variety of referential types includes the concept in the respective propositional functions. These functions make the repetitions of *grammar* quite different from the repetitions of *language*, which, as was pointed out above, broadens the scope of the objects under review in the article. The repetitions of *grammar* include it in various argumentative functions and thus place it in the centre of developing the author's thesis.

Repetitions of *study*

So far we have seen recurring forms of an adjective and two types of repetitions of nouns, one of which changes the intension with each occurrence, the other recurs in varying referential types. The repetitions of the word *study,* for their part, are characterised by frequent reference to the concept *the study of language*; an even longer string is established–*contribution to the study of language*. The reference varies–generic, specific, substitutions, restructuring. There even is one case of the noun used in a different sense:

> *can be reached from a **study** of his materials beyond the*

However, such a shift is rather an exception to the rule. Mostly, authors do not use a word in homophonic realisations.

Repetitions of *problem*

A different type of repetition is exemplified by the lexeme *problem*. As can be seen from the identifying phrases below, two different concepts are picked up:

__problem__ of acquisition of knowledge
proposal to deal with the __problem__ of acquisition of knowledge of
i have been describing the __problem__ of acquisition of knowledge of
the exact nature of the __problem__ of acquisition of knowledge

would then face the __problem__ of explaining how the pre-linguistic
ite senseless to raise the __problem__ of explaining the evolution of
ted that there exists a __problem__ of explaining

Several cases of phrases where *problem* appears as a substitute are found in the text. As can be seen from the examples below, a different problem is named each time:

speculation, however, has no bearing one way or another on those aspects of the __problem__ of mind that can be sensibly pursued. It seems to me that these aspects

structure of the visual cortex. No one who has given any serious thought to the __problem__ of formalising inductive procedures or "heuristic methods" is likely to

d as the actual theory of the language in question. I have been describing the __problem__ of acquisition of knowledge of language in terms that are more familiar

is possible. Peirce, to my knowledge, is original and unique in stressing the __problem__ of studying the rules that limit the class of possible theories. Of cour

I doubt that it has been fully appreciated to what extent this complicates the __problem__ of accounting for language acquisition. Formally speaking, the learner m

or suggests a plausible alternative approach, with empirical content, to the __problem__ of acquisition of knowledge. Assuming the rough accuracy of conclusions

model for investigation of other cultural and social systems. In general, the __problem__ of extending concepts of linguistic structure to other cognitive systems

estrictive. The third sub-task, then, is to study what we might think of as the __problem__ of "confirmation" — in this context, the problem of what relation must

what we might think of as the __problem__ of "confirmation" — in this context, the __problem__ of what relation must hold between a potential grammar and a set of data

the __problem__ of learning, but will rather offer an incorrect solution to this __problem__. The issue is an empirical one of truth or falsity, not a methodological

> *empirical assumption. As I have argued earlier, a non-dogmatic*
> *approach to this **problem** can be pursued, without reliance on unargued*
> *assumptions of this sort —*
> *earning is how this invention of grammar can take place. Putnam does*
> *face this **problem** and suggests that there might be "general multipurpose learning*
> *strategi*
> *age in these domains. No one, to my knowledge, has devoted more*
> *thought to this **problem** than Lévi-Strauss. For example, his recent book on the*
> *categories of pri*

This type of noun is characterised by the flexibility of its conceptual load. What is actually included with each repetition depends on the immediate context, rather than on some general scope particular to the lexical item.

Repetitions of *system* and *systems*

Finally, we take a look at two forms of a lexeme which occur with equal frequency – *system* and *systems*. The latter names two types of objects: symbolic systems and communication systems, while the former applies to both and a few other referents in phrases of the type we called **substitute**.

The plural noun appears in concordances like these:

> *that there exists a problem of explaining the "evolution" of human*
> *language from **systems** of animal communication. However, a careful look at*
> *recent studies of an*
> *There have been some attempts to study the structure of other, language-like*
> ***systems** — the study of kinship systems and folk taxonomies comes to mind, for ex*
> *what human language is, we find no striking similarity to animal*
> *communication **systems**. There is nothing useful to be said about behaviour or*
> *thought at the le*
> *assumption that there is an evolutionary development of language from*
> *simpler **systems** of the sort that one discovers in other organisms. Popper argues*
> *that th*
> *for a moment. The assumption that human language evolved from more*
> *primitive **systems** is developed in an interesting way by Karl Popper in his recently*
> *published*
> *of the universal features in a fundamental way, but it is the properties of*
> *the **systems** of rules, it seems to me, that really shed light on the specific nature*
> *Lévi-Strauss occasionally alludes, becomes meaningful only when one*
> *considers **systems** of rules with infinite generative capacity. There is nothing to be*
> *said*
> *were present in some form in these already acquired prelinguistic*
> *"symbolic **systems**." But since there is not the slightest reason to believe that this is*

*evidence that the mind is simpler in its innate structure than other biological **systems**, just as it would be mere dogmatism to insist that the mind's organisation*

The uses are mainly generic. The attributes are in the semantic range of communication systems to systems of language rules.

The singular noun can be found in concordances exemplified below:

Citation form:
*man's case, the argument is based entirely on a vague use of the term "symbolic **system**," and it collapses as soon as we attempt to give this term a precise mean*

Identifying forms – different referents:
*those of Lord Herbert and Descartes, both of whom took for granted that the **system** of innate ideas and principles would not function unless appropriate stimulus*
*e at the moment about the general properties of the underlying phrase structure **system** for natural languages; the dispute is not in the least resolved by the ex*
*that the a priori is due to hereditary differentiations of the central nervous **system** which have become characteristic of the species, producing hereditary dis*
*properties of the physical world is based on innate organisation of the neural **system**. In some cases at least, these built-in structures will degenerate unles*

Definite generic forms
*"simplest possible" one would have to demonstrate that the "optimal" computing **system** would take a string of symbols as input and determine its surface structure*

Indefinite Generic forms:
*conclusions involves a false assumption. From the fact that a phrase structure **system** contains proper names one can conclude almost nothing about its other cat*

Identifying indefinite uses:
*that "acquisition of an initial language is acquisition of a secondary symbolic **system**" and is quite on a par with normal second-language acquisition. The prima systems provide the "algorithms which are 'simplest' for virtually any computing **system**," hence also "for naturally evolved 'computing systems' "; and that there invariant through long historical eras. Furthermore, we discover a substantial **system** of principles that do not vary among languages that are, as far as we know ws," nevertheless held firmly that underlying any human language we will find a **system** that is universal, that simply expresses man's unique intellectual attrib seem tenable today, it is reasonable to suppose that a generative grammar is a **system** of many hundreds of rules of several different types, organised in accord*

*tence — knowledge of a language — as an abstract **system** underlying behaviour,
a system constituted by rules that interact to determine the form and intrinsic me
we must regard linguistic competence — knowledge of a language — as
an abstract **system** underlying behaviour, a system constituted by rules that
interact to determine*

Substitutes:
*ures or "heuristic methods" is likely to set much store by the hope that
such a **system** as a generative grammar can be constructed by methods of any
generality.*
*ation, is common to all languages. There is no a priori "naturalness" to
such a **system**, any more than there is to the detailed structure of the visual cortex.
N*

As can be seen, the singular phrases differ both in their referents, that is,
name different concepts, and in the type of referentiality. Some uses are
generic, others–specific and all of them are included in different
propositional actions: generalising, giving examples, referring back to
previous uses, picking up iterative referents etc. The discussion revolves
around the issue of searching for and interpreting systems for
communicating ideas. Unlike the repetitions of *language*, these ones
present an evolving argument. On most occasions the repeated noun forms
part of naming complexes with other lexemes, where the meaning is the
result of pulling together intensions from the component parts.

Conclusion: patterns

Therefore, four types of repeated items are established:

1.   Repetitions outlining the conceptual field of investigation. They
maintain reference to selected concepts under discussion in the article. The
term that can name this type is **conceptual/analytical chain**. An example
is the chain of repetitions of *language*. The repetitions in this group occur
with changes in the intension, which allows the author to broaden the
scope of the discussion.
2.   Repetitions describing the analytical framework within which answers
are sought to the research question. We can adopt the term **illustrative
chain**. This type is exemplified by the repetitions of the forms *system* and
systems. These chains are immediately visible in the frequency list of
repetitions, because they include both the plural and the singular form in
roughly equal numbers. The referential types differ, which allows the
author to include them in developing his argument through a range of
propositional functions.