

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

Elena Tarasheva¹

Abstract

The presentation reports the findings of a PhD study of lexical ties in text. Unlike other studies, the present one focusses on the repetition of word forms, exploring the word form as an immediately obvious unit of analysis. Texts from three genres: short stories, research articles and political speeches in 2 languages (Bulgarian and English) are studied to establish:

1. Is it true that repetitions are avoided in written texts?
2. Do repetitions of word forms group into recurrent structures?

The method of study is to compare the form, intension and referent of each occurrence of a repeated word form in a text, thus looking for recurring structures.

1. Research framework

The question which lexical relationships have a text-building function tends to trigger complicated procedures for admittance to the category of linking lexis. Hoey (1991: 58) established a flowchart for this purpose, Morris and Hirst (2002) traced thesaural indices, Viehweger (1976) was criticized for positing a widely associative principle for his thematic fields which hardly prevents any items from featuring as linking. Therefore, it is usually tacitly assumed that any lexically related words have a linking function as long as they can be shown to share a referent in the concrete text. Even if we ignore the fact that the establishing procedures are time-consuming and difficult, in the end, quite a few seemingly related items happen not to function cohesively in a text for various reasons. Moreover, from the theoretical point of view, should co-reference and lexical links present different text-building mechanisms, there must be a way to separate the two.

For the attempt to study lexical links in isolation from co-referring terms, a suitable unit of analysis appears to be the word form, defined as ‘material between two spaces’ (Sinclair, 1991: 28). The word form is indisputably lexical in essence, it is also discrete and easy to establish by electronic instruments. Should the repetitions of word forms create recurring structures in texts, then they can be seen as the basis of lexical ties.

The hypothesis that the repetitions of word forms build the structure of a text can be tested by exploring the repetitions in concrete texts and checking whether they follow a specific pattern. The parameters of each occurrence of a word form can be formalised using Petofi’s semiotic pyramid (1985), which represents the lexicon item as a unity of form, sense (intension) and referent (extension).

¹ Department of Linguistics, New Bulgarian University
e-mail: etarasheva@nbu.bg

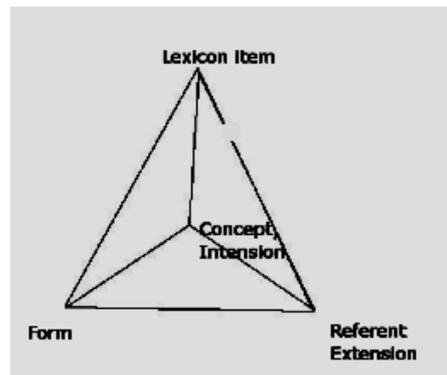


Figure 1: The lexicon Item as presented by the semiotic Pyramid

This study assumes that the apex ‘form’ coincides for the repetitions of word forms. The analysis is expected to establish whether the intensions and extensions coincide or differ for subsequent occurrences of the repeated item. Representing ‘lexicon item’ and ‘form’ as different apices in the model proceeds from the assumption that the word form and not the lexeme is the unit of analysis. This approach also allows positing that a lexeme can be represented by more than one chain of repeated word forms – e.g. one chain for the plural and one - for the singular.

It is a tradition to consider some uses of lexicon items as referential because they name a specific entity, while others are non-referential because they do not correspond to anything in the real world. To acknowledge this distinction in my analysis, I adopt Thrane’s (1980:39) classification of identification in one of two possible spaces: the real world (spatiotemporal identification) or in the generic (categorical) space. When identifying in the generic space, the named entity picks up either a typical representative of the class, known in traditional grammars as ‘a generic name’, or any representative of the class – a process usually referred to as ‘specification’, or a non-existent representative. In my model such identifications do not realise the apex ‘referent’. The apex ‘intension’, for its part, is defined by Petofi (*ibid.*) as including all the knowledge about the named entity. Consequently, I consider that each occurrence picks up a specific meaning of the lexicon item. If all the meanings enumerated in a dictionary present the full set, each use employs part of the set. Further, as has been described in literature (Sax 1975, Nunberg 1978, Green 1989, among others), pragmatic shifts take place in examples such as *I drank the whole glass*, where ‘the glass’ refers to the content inside. Alternatively, metaphoric shifts of the intensional content occur in uses such as *the heart of the forest* or *a window of opportunity*. Thus a lexicon item used in speech can name a referent from the real world or from the generic space; additionally, the speaker can control the intensional content by selecting from the meanings of the word, or shifting them metaphorically or pragmatically. All the repetitions of a word form in a text can pick up the same referent or different referential and intensional projections.

2. Corpora

Given that patterns of repetitions can vary according to genre, for the purposes of my study corpora of texts from three genres were created: short stories, research articles and political speeches. To encompass peculiarities which can be language-specific,

each genre included texts in 2 languages: Bulgarian and English. Recurring patterns were sought across the genres and languages in the study.

- 9 short stories in English by established authors – 61 125 running words
- 9 short stories in Bulgarian by established authors – 25 233 running words
- 8 research articles in English by experienced researchers – 77 583 running words
- 8 research articles in Bulgarian by experienced researchers – 19 297 running words
- 10 political speeches in English by established politicians – 26 860 running words
- 10 political speeches in Bulgarian by established politicians 11 571

In view of the fact that experienced writers could be expected to create more coherent texts than inexperienced ones, supplementary corpora were created from texts written by students. The repetition patterns were qualitatively and quantitatively compared.

- 4 short stories in English – 6,657 word forms
- 4 short stories in Bulgarian – 5,841 word forms
- 7 research papers in English – 7,951 word forms

As the repetitions were studied within the unity of each specific text, a normalisation procedure was needed to overcome the disparities in the size of the texts. Using Biber's (2000: 263) approach, the number of repetitions was presented per one thousand words. Thus an index was established for each text. For each genre and language the highest and lowest indices were compared.

3. Procedure

The repetitions of word forms in each concrete text from the corpora were established by a concordancer. Only notional words were considered (although function classes, such as modal verbs and pronouns, were taken into consideration for the political speeches). All the occurrences of a word form were considered to form a chain. For each occurrence of a repeated word form I answered the following questions:

1. In what complex does the repeated word form occur: definite, indefinite or zero-marked? Is it plural or singular?
2. What object is named by the referring complex? Do all the occurrences of the word form name the same referent?
3. Is the referent identified in the categorial or the temporal space (referential projection)?
4. Are there pre- or post-modifying expressions which present parts of the referent or other semantic relations (restructuring)?
5. Does each occurrence of a repeated word form evoke the same intension or is there some metaphoric, pragmatic or another shift?

The analysis was carried out using the Linguist's Workbench developed by Maria Stambolieva (1996), which works with texts in English and Bulgarian.

Statistical data were obtained through the programme Build. The programme Concordance was used to study the linguistic context of each repetition.

Each chain could be expected to form a pattern of representations of referents, definite or indefinite phrases and pre- or post-modifying complexes. This would be considered a recurring regularity. The regularities could depend on the genre, the semantics of the repeated lexicon item *etc.*; they could obtain across the two languages or be specific for one of them.

4. Analysis of the data

4.1 Quantitative data – indices

Table 1 presents the indices of repetitions in the 3 genres and the two languages:

	Lowest index (per 1000)	Highest index	Lowest index	Highest index	Lowest index	Highest index
English	691	841	618	778	402	685
Bulgarian	484	617	474	642	434	610
	Research Articles		Short stories		Political speeches	

Table 1: The indices of repetitions in the three genres

The data show that even if repetitions were avoided, they amount to at least 600 per one thousand words in each text in English, and 400 per one thousand words - in Bulgarian. With approximately one half of the words repeated in each text, repetitions of word forms can be considered a quantitatively significant factor in texts. The only genre where repetition appears to be avoided, surprisingly, is rhetoric, where the index for the English texts drops to the unprecedented 400 per one thousand words. A feasible explanation of this phenomenon is that in oral speech supra-segmental features, such as stress and intonation, take over the linking functions of repetitions, which keeps the index significantly lower than in the other genres.

The number of repetitions is significantly higher in English because in Bulgarian the definiteness marker is a gender specific bound morpheme and it is applicable to nouns, adjectives and numerals, which makes the respective grammatical paradigms considerably more sizeable than the English paradigms for those morphological classes. For the Bulgarian verbs inflectional, tense and aspect markers make the paradigm even lengthier than for the substantives. For example, while in English each noun has two forms – for the singular and plural, in Bulgarian a noun has two forms for number and three – for definiteness, if it is masculine.

From Table 1 it also might look as if the pattern of repetitions in texts written in Bulgarian does not vary significantly across the genres. This has been explained with the role of editors and the general perception that repetitions should be avoided as much as possible - a tendency much regretted by translators from English and German.

It is also the case that inexperienced writers repeat significantly less often than experienced authors. For example, in a research article of 1333 running words, an

experienced writer repeats the term from the title of the article *parallelism* 22 times; in an article of 1121 words a student uses the term from the title *identity* 7 times. Further differences exist in the respective referential patterns. Therefore, repetitions do not decrease the quality of a text – on the contrary, they allow the author to present more aspects of a notion while referring to its identity by the same unambiguous term.

4.2 Qualitative data: chain typology

The occurrences of repeated word forms reveal a recurring pattern of referential presentations. For example, general substitutes, such as *man*, *woman*, *child* in short stories appear in two types of uses: (1) classifying phrases, where no referent is picked up; and (2) referential phrases, where they name different characters who are not given names in the short stories. In the latter case the referents differ for each occurrence. For example, *man* occurs in this concordance in one of the stories:

it was the very	man	whom albert had described lying
she looked up at the	man	sitting opposite her to see
towards that no doubt excellent	man	the author decided not to
her agent a round faced	man	who wore glasses so strong
of him as a little	man	
he was a very young	man	not thirty but he had

Example 1: Concordance of general substitutes

All the short stories in English and Bulgarian contain chains of general substitutes and they have the same type of reference, albeit in different quantities for each case.

Parts of the setting are described by repetitions included in referential noun phrases, mainly with the definite article, naming a different referent each time, as can be seen from Example 2:

when he comes to the	door	he always asks is there
a moment he opened the	door	in his hesitating way and
miss warren at the	door	in order not to interrupt
the	door	slammed behind her
the	door	was opened by a scraggy
house he went to the	door	as they entered the drawing-
a second then held the	door	wide open
the room and closed the	door	
afternoon presenting myself at the	door	instead of asking the maid
the carriage is at the	door	
bulfinch opened the	door	and asked her if she
a	door	opened as she reached the
a	door	at the back covered with
albert's key in the front	door	they realised with one accord
a small poky bedroom next	door	to the kitchen for mr

Example 2: Concordance of parts of the setting

In research articles the material under study is named by a lexeme which always appears in two chains – of the plural and the singular word forms. The two are approximately equal in number. The analytical tool is always named by a phrase

which occurs with numerous pre-modifying restructuring phrases; in Bulgarian the topic chain always splits in two: the chains of the lexeme for the definite and for the indefinite.

Thus words naming particular types of entities appear in a configuration of referential representations reflecting the functioning of the entity in the genre. On the basis of these features, the chains of repetitions of word forms were classified as is shown in Table 2:

Political Speeches	Short Stories	Research Articles
	Topic chain	Topic chain
	Characters	
Nominative chains	Parts of the Body	Analytic chains
	Surroundings	Illustration
Block repetitions	Verbs	
	General Substitutes	Research Vocabulary
	Temporal Chains	
Chains of Attributes	Chains of Attributes	Thematic fields
Rhetoric Chains		
Deictic Chains		
Modality Chains		

Table 2: Chain typology

For the short stories and the research articles the types of chains seem to include the essential parts of the respective type of exposition. If a short story is about the characters, their attributes, actions in the particular setting, a research article is an account of a study with its particular analytical tools, specific illustration, including certain thematic fields and using a range of research vocabulary. Thus the repeated word forms create the schemata of the respective genres. Moreover, only student writing is found to exclude types of chains from the established taxonomy, or to include chains from another genre. For example, a research article containing a temporal chain causes the article to sound like fiction. Alternatively, a research article without chains of research vocabulary reveals little about the research process. Likewise, too many topic chains show that the student attempts to tackle more issues than can be approached in one study, or a lack of analytical chains is indicative that the author is just speculating but not analyzing the issue by appropriate research methods. Therefore, a good research article or short story contains all the types of chains established with the present study and no chains alien to its genre. Conversely, to answer the research question of the present study, the repetition chains can be seen as creating the genre skeleton of a text.

A slightly different typology emerges from the political speeches. The taxonomy of the chains reflects a whole communicative situation, rather than the genre framework, as was demonstrated for the written genres in the corpus. The chains feature the participants in their specific roles, the modality and the style of the speaker. The pronominal chains reveal a set of referents specific for each speaker. For

example, G. Bush uses ‘we’ (the rich countries) in confrontation to ‘they’ (the third-world countries), John Kennedy in his inaugural speech includes his ‘worthy predecessors’ in the reference of ‘we’, Churchill, for his part, refers to ‘we’ as the anti-fascist coalition *etc.* The modality chains repeat the exponent of a dominant grammatical category, thus setting the mood of the speech. For example, in Blair’s speech justifying the intervention in Iraq the verb *have* is repeated in combination with **would** to express unfulfilled options:

threat to britain we would	have	taken action in september we
as they do would	have	taken heart
perhaps we would	have	found different ways of reducing
all of this nothing would	have	happened
have killed or they would	have	rejoiced in it
rogue states and wmd would	have	been shown to be hollow
inspectors would	have	stayed but only the utter
he would	have	strung the inspectors out and
of the iraqi people would	have	continued unabated and reinforced
and its partners saddam would	have	cooperated more
faced with this decision would	have	gone the other way for
its democracy crushed it would	have	been a long time
but if they could	have	killed or they would have
september we would not	have	gone to the un

Example 3: The Concordance of *have* in Blair’s speech on Iraq

For his own purposes, Edward VIII repeats *have* in a perfective sense, thereby setting the modality of a report of the current situation:

during these hard days i	have	been comforted by her majesty
emperor and now that i	have	been succeeded by my
i occupied the throne i	have	been treated with the
know that the decision i	have	made has been mine and
	have	made this the most serious
baldwin the prime minister	have	always treated me with
and lately as king i	have	for twenty-five years tried to
i tell you that i	have	found it impossible to carry
all know the reasons which	have	impelled me to renounce the
of the people wherever i	have	lived or journeyed throughout
i	have	never wanted to withhold anything

Example 4: The Concordance of *have* in Edward VIII’s abdication speech

Other types of chains project the specifics of the speaker’s style. Rhetoric chains typically contain items such as *all, both, not etc.*, thus revealing the methods in which a speaker generalises or negates. The type of repetition where a whole complex reappears, called here block repetitions, emphasizes the constructed nature of political discourse. When a nominative phrase is repeated co-referentially, and the second mention reiterates the previous one entirely, it creates the impression of anaphora, as if the referent’s existence is taken for granted. The corpora for this study include examples of this in Blair’s speech about Iraq - *an imminent threat*, in president Bush’s speech about the problems of the third world countries - *the democratic community*, in Churchill’s speech - *the iron wall* - *etc.* All of the respective referents have been contested on various occasions, but their representation in the speeches tries to prevent challenges by the security of the anaphoric reference. The coming into existence is simulated by the anaphoric block repetition.

These peculiarities show that the repetitions of word forms in political speeches are instrumental in creating the framework of a communicative situation where a person addresses an audience with a message. Compared to the strictly written genres in the corpora, it becomes obvious that lexical repetitions create a skeleton known from psychological research (Schank and Abelson 1977) as the frame of the situation. The fact that the repetitions of word forms are characterised by this feature gives them a text formative status.

4.3 The role of the intension and extension

Within the chains of repetitions – which were established as configurations of recurring referential representations – the majority of the occurrences set off either the intension, or the extension of the lexicon item. In the short stories, the referents loom large, while in research articles the intensions are emphasized. While the repeated word forms in fiction reinforce the identity of a named character or place, in research articles the intension is under construction. Each successive occurrence repeats the intension with modifications thus enriching the intensional content of each concept and creating a constellation of meanings typical of the article. Example 5. shows the shifts in the intensional content of the repeated word form *knowledge*.

body of knowledge with western	knowledge	has been one of the
as for the transmission of	knowledge	higher education is still largely
de-revolutionising	knowledge	replacing the communist body of
the producers of symbols of	knowledge	and wisdom
need for graduates with different	knowledge	and skills but by the
the countries moving towards the	knowledge	society
not motivated to update their	knowledge	

Example 5: The Changing intensional content of *knowledge* in a research article

As can be seen from the concordance, successive occurrences imply both the fact of knowing and the sum of what is known, as well as rare meanings, such as ‘qualifications for knowledge’ and ‘the traditional body of knowledge’. As a result, the message of the article is concerned with these meanings rather than with a different set. Therefore, each author evokes a certain range of meanings to create an intensional content specific for their article.

In fiction the intension of the named entities is more or less stable throughout the text. Authors, for their part, tend to play with the referential types. The main part of a plot in a short story is characterized by varying reference, as can be seen from Example 6:

GENERIC USES:

a man doesn't run away with his	cook	without making his wife
ridiculous		
you any harm but a	cook	would finish you

IDENTIFYING USES:

molire and his	cook	said miss waterford
let albert keep his	cook	
for his affections with my	cook	
my	cook	

albert has eloped with the	cook
bringing it in and the	cook got all upset about it
floor and she recognised her	cook

you will find the new	cook satisfactory
you mean by the new	cook '
my menus with the new	cook ma'am

CLASSIFYING USE:

treasure that woman a remarkable	cook and she gets better and
----------------------------------	-------------------------------------

Example 6: The Concordance of the main part of the plot with changing referential projections

When an author uses a word in generic reference, he or she can make generalizations; if the word is used referentially, various representatives of the class can be named; the classifying uses of words are instrumental in attributing qualifications. Therefore, different referential types include the named entity into the mental processes of generalizing, giving examples, speculating *etc.*, which are part of the coherence structure of the text. Thus repeating a word form with different referential types makes it part of the coherence structure. Conversely, for corpus analysis, varying forms of reference signal significance for the plot.

Even adjectives take part in the plot by a change of reference:

AN IMAGINARY REPRESENTATIVE:

i thought if i was	lucky it might stop whispering
i thought if i was	lucky
mother to know i was	lucky said the boy
but if you're	lucky you will always get
if you're	lucky you have money
lucky and you must be	lucky because it was your ten
do you think i'm	lucky mother

POSITIVELY CLASSIFIED:

he said stoutly i'm a	lucky person
i am	lucky
bassett and i are	lucky and you must be lucky
i call that	lucky don't you mother
better to be born	lucky than rich

NEGATIVELY CLASSIFIED:

and is father not	lucky
so i thought you were	lucky
and aren't you	lucky either mother

GENERIC:

knows why one person is	lucky and another unlucky
-------------------------	----------------------------------

Example 7: The Concordance of adjectives with differing referential projections

The various analyses of coreference do not take into account non-referential items, such as adjectives and verbs. The chain in example 7. would not feature in a study of co-referential chains, which deprives the study of a significant detail in the story.

Conversely, chains for parts of the human body in the short stories - occur in identifying referential uses only:

end the sneering yellow **faces** of young men that met
 only of the watchful yellow **faces** behind
 the lock-ups the grey cowed **faces** of the long-term convicts
 at the sea of yellow **faces** above the garish clothes-faces
 the will of those yellow**faces** behind

Example 8: The concordance of unchanging referential projections

Such items act as background to the plot. The coherence structure of the story recognises them as a descriptive backdrop, unlike items with changing referential projections. Similar ancillary behaviour is observed with components of the setting, such as doors, windows, chairs *etc.* No generic uses can be found in the repetition chains naming such entities. However, variety is provided by metaphoric shifts, exemplified here by *the sea of faces*, and the restructuring of the referent by the pre-modifying phrase *the will of*.

The ties established by the repetition of adjectives link referential and non-referential uses by their attributive function:

the wing and delivered a **military** command of some sort in
 of spirit but of no **military** experience was made captain
 as he was concerned this **military** expedition of ours was simply
 sort in a large **military** fashion nobody was surprised at
 the time and admired the **military** ingenuity of the device but
 had made our first **military** movement and it was a
 was not entirely destitute of **military** resources and i did
 not right and proper and **military** for me to have an
 full of uncertainties about everything **military** so i let the matter

Example 9: The linking role of adjectives

As is seen from the example, nouns such as *command*, *experience*, *expedition*, *movement* and *resources* turn into a thematic list by virtue of the connection with the adjective *military*. Moreover, the chain names a central motive in the text.

The researched material for this study gives ample evidence that repetitions in short stories maintain the reference to the referential set of a lexeme, though not to one single representative in all the occurrences. Semantic classes such as general substitutes, parts of the human body and of the setting easily switch reference among various representatives of the class. Identifying the referent in the two possible spaces – the generic and the spatiotemporal – highlights the entity through inclusion in the coherence structure; conversely, restricting the referent to the spatiotemporal domain places the entity in the background. It can also be said that the tendency is to vary the method of naming by including restructuring phrases, metaphoric and pragmatic shifts when repeating a word form.

In research articles co-referentiality plays an insignificant part as a linking device. Most of the terms used there are non-referential, therefore among the repetitions intensional links can only exist. Isolated instances of referential use are found in the illustrative chains, an example of which is shown in Example 10.

unequivocally that when a spoken **stimulus** is processed some of its
 thus for example the **stimulus** s _ ba was followed
 down the list by the **stimulus** seba these two words differ
 headphones every seconds for each **stimulus** the participant had to
 decide

after the offset of the	stimulus	response time was measured
from		
from the onset of the	stimulus	
created four counterbalanced lists of	stimuli	in the following way in
vs non-words and list of	stimuli	
of the instructions played the	stimuli	off the hard disk and
method material all the	stimuli	used in the experiment were
	stimuli	were presented through
		headphones every
a catalan speaker reading those	stimuli	at a pace of one LR

Example 10: Illustrative chains in research articles

Four repetitions name concrete referents, one evokes iterative representatives (*each*), two are generic names, the rest restructure the referent for naming aspects, such as the onset of the stimulus, stimulus-response time and a list of stimuli. Therefore, what all the occurrences of the repeated word form have in common is the intension. Example 10. also demonstrates the typical feature of illustrative chains to be represented by an equal number of singular and plural items as was mentioned in 4.2. above.

Adding new information with each new occurrence of the word form is seen in research articles even more often than in fiction. Apart from the shift in the intensional content demonstrated in Example 5., research articles show the tendency of restructuring the referent. Different pre-modifying phrases and attributes evoke aspects of the named referent, which process is called here ‘restructuring’.

overall the	reform	of east european higher
education		
constrained rather than supported the	reform	of the content of higher
and eventually reversing the initial	reform	initiatives
the whole	reform	discourse reaches a
magnificent level		
romelaer describes much better the	reform	processes in east european
higher		
the goal of such	reform	ideas was to restore the
between and on university management	reform	
as well as external	reform	agents a full conservative
restoration		
is a need for a	reform	as disclosing such an
in east european higher education	reform	documents
based on various countries	reform	efforts the report identifies
the		
shed social institutions has produced	reform	environments that are rather
similar		
europa the initial requests to	reform	higher education systems were

Example 11: The topic chain in a research article

The author restructures the repetitions of the item *reform* by talking about the reform in East European Universities, about various reform initiatives, by analysing the reform discourse and ideas *etc.* By tracing the restructuring, one can get a clear idea of the message. This can be seen as the basis of a summarising mechanism whose algorithm focusses on the types of repetition chains established here.

Example 11 also demonstrates the equal number of definite and indefinite occurrences in topic chains, immediately obvious in Bulgarian but seen in the

concordance of the terms in English. Therefore, the type of chain can be easily established by numerical criteria.

Another argument reinforcing the significance of restructuring can be found in student writing. The repetition chains there are characterised by imprecise and repetitive reference, restructuring does not unfold a range of aspects through which the object of the study is analysed. Below is the concordance of *identity* – the key issue in the topic of the student’s essay:

modernism in american writing back to the issue of bridge continental ideological and mainly name and as if no being american or not and yet so indisputably american	identity is not a modern issue identity issues it is just a identity identity gaps and turn the identity identity is a matter of choice identity
--	---

Example 12: The topic chain in a student’s research paper

Moreover, restructuring in the experienced researchers’ papers takes place in the chains in a special layered fashion, as can be seen from Example 13:

transition 70. cohesive 53. type 17. types 17. states 9. transitions 70. semantic 72. similarity 45. distance 17.	relationships 12. measure 19. measures 11. cohesion 57. inferable 70. centers 47. cb 56. discourse 69. entities 43. elements 18.
---	---

Example 13: The layered restructuring of concepts in research articles

The arguments in the discussion can be seen through the schematic representation above: the subject of cohesive transitions is discussed by taking a look at transition types and transition states. Semantic relationships such as similarity and distance are measured by specific measures.

Therefore, research articles develop the intension of a few selected concepts by shifts in the meaning and by restructuring the repetitions to name selected aspects of the respective notions.

In conclusion, the extension and intension of the repeated items have a special role to play in the chains of repeated word forms. On the one hand, they conform to the specifics of the genre, on the other they include the item in the coherence structure of the text. The chains in the corpora for this study reveal recurring patterns, thus confirming the text formative function of the repetitions of word forms.

5. The chain taxonomy and the content of the texts: summarising

So far, the study established that repetitions group into types of chains building the skeleton of each genre. Within each chain a specific configuration of generalisations, classifications and identifications are encountered. Having in mind the finding that the

restructuring of the main concept outlines the content of a research article, I tried to compare the contents of articles to what is projected through the chain typology. In a short popular article by Chomsky, the chain of repetitions of *universality* shows the specifics of a topic chain:

opens new paths to studying	universality in language
extends to my topic here	universality in language and human rights
things to say about	universality in language and about universality
say a few words about	universality in language and in human
to begin with what about	universality in language
universality in language and about	universality in human rights but
from its relation to the	universality of human rights
directly to the issue of	universality of human rights
the principle of	universality arises in other connections too
which bear quite directly on	universality in language insofar as

Example 14: The topic chain

The definite occurrences are almost equal in number to the indefinite ones; restructuring is established, therefore the chain can be considered a topic chain. Generic reference dominates. Two types of universality are discussed – the universality of language and the universality of human rights. The topic is restructured by the phrases *the issue of ~*, *in relation to ~*, *a few words about~*.

The analytical chains – consisting of equal numbers of singular and plural forms – are two: *language/languages* and *judgment/judgments*. They collocate strongly with respectively *internal* and *moral*. The thematic fields are *universal grammar*, *universal human rights*, *cognitive revolution*, *general principles*, *human language/ rights*, *world bank/court*. The scientific vocabulary includes items, such as *systems*, *range*, *case*, *properties*, *inquiry*.

Therefore, the summary would be:

The universality of language and human rights is discussed in terms of the universality of internal language and moral judgments. Arguments from the cognitive revolution and universal grammar are brought alongside cases from the practice of the world bank and courts. General principles are sought quoting different inquiries and commenting on the properties of language and human rights.

Compare this with Chomsky’s short abstracts at the beginning of his article:

Thirty-five years ago I agreed, in a weak moment, to give a talk with the title “Language and Freedom.” When the time came to think about it, I realized that I might have something to say about language and about freedom, but the word “and” was posing a serious problem. There is a possible strand that connects language and freedom, and there is an interesting history of speculation about it, but in substance it is pretty thin. The same problem extends to my topic here, “universality in language and human rights.” There are useful things to say about universality in language and about universality in human rights, but that troublesome connective raises difficulties.

The only way to proceed, as far as I can see, is to say a few words about universality in language, and in human rights, with barely a hint about the possible connections, a problem still very much on the horizon of inquiry.

The comparison shows that the taxonomy of chains leads to correct genre identification and summarizes the contents adequately. Therefore, the chains of lexical repetitions apart from creating the genre skeleton, reflect the content of the text. Used in accordance with their classification established with the present study, they can lead to summarizing.

6. Conclusions

The present research has demonstrated that the study of repetitions of word forms leads to establishing recurrent structures which form genre-specific chains. The taxonomy of the chains creates a skeleton projecting the specifics of the communicative situation in each genre; it also recreates the content of the text correctly. Through a particular configuration of generic, specific and other referential projections, the repetitions allow the author to make generalisations, exemplify, infer and perform other mental operations with the entity named by the repeated term. Through pre- and post-modifying expressions the authors shed light on different aspects of the concept in focus. The repetition chains encapsulate the meaning of texts, as well as the genre framework. Given the enormity of the phenomenon – more than sixty percent in English and forty percent in Bulgarian – repetitions are a significant factor in creating texts.

The outcomes of my research can lead to new methods of summarising and assessment of student writing.

References

- Douglas, B., S. Conrad and R. Reppen (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP
- Green, G.M. (1989) *Pragmatics and Natural Language Understanding*. Hillsdale, NJ: Laurence Erlbaum
- Halliday, M. and R. Hasan (1976) *Cohesion in English*. London: Longman
- Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: OUP
- Morris, J. and G. Hirst (1991), "Lexical cohesion, the thesaurus, and the structure of text" in: *Computational linguistics*, 17(1), 21–48
- Nunberg, G. (1978), *The Pragmatics of Reference*. Bloomington: Indiana University Linguistics Club Publications
- Petofi, J. (1985) *Lexicon*, in van Dijk, T.A. (ed.) *Handbook of Discourse Analysis*, pp 87-100. London: Academic Press
- Sacks, H. (1975) *Everyone Has to Lie in*: Jaworski, A. and N. Coupland (eds) (1999) *The Discourse Reader*, 167–82. London and New York: Routledge
- Schank, R., and R. Abelson (1977) *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford : OUP
- Stambolieva, M. (1996) "A Linguist's Workbench" *Linguistics* 28, 293–301
- Thrane, T. (1980) *Referential-semantic analysis. Aspects of a theory of linguistic reference*. Cambridge: CUP

Viehweger, D. (1976) Лингвистика текста в исследованиях ученых ГДР in
Золотова , Г. А. (ed.) Синтаксис текста, pp 314–23. Moscow: Nauka
(Russian translation of ‘Advances in Text Linguistics’)