CHAPTER TWO

REPETITIONS IN RESEARCH ARTICLES

Repetitions of Word Forms in Texts: An Approach to Establishing Text Structure
Author: Elena Tarasheva
Date Of Publication: Feb 2011
Isbn13: 978-1-4438-2662-4
Isbn: 1-4438-2662-6

http://www.c-s-p.org/Flyers/Repetitions-of-Word-Forms-in-Texts--An-Approach-to-Establishing-Text-Structure1-4438-2662-6.htm

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

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 a noun, 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. Concepts outlining the conceptual field of investigation. The repetitions 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. Concepts 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.

3. General research vocabulary. This helps refer to concepts and operations for further discussion and qualification. The term to name the category can be general research vocabulary. Such a repetition chain is that of the word *problem.* Substitute forms, shifting reference and predominant specific-ness characterise this type of chain.
4. Finally, a chain seems to deserve the name attributive repetitions. They present adjectives which restrict the reference to selected areas. This is the example of the adjective *human*.

Each of the four types contributes to the development of the article in its specific way. In each case the reference of the phrases in which the repeated items are included follow a different pattern.

Part Two: Does the pattern occur in other articles?

The fact that such a pattern seems to exist in one article, however, is not informative enough in view of the role of repetitions. Is it the case that such repetitions occur in other research articles? The corpus of 8 articles was studied for repetitions belonging to the types established above. They all contain the four types of repetition chains, as can be seen from table 2.1:

Table 2.1. Types of chains in the research articles

Attributive	**Conceptual /Analytical**	**Illustrative**	**General research vocabulary**
lexical 36. identical 7. syntactic 7. japanese 6. ultrasonic 6. intersentential 4. technical 4. lexico-semantic 3.	parallelism 22. repetition 9. beginning 7. device 5. wave 5. amplification 4. dependency 4. distance 4. information 4. traveling-wave 4. kanji 3. keywords 3. substitutes 3. ???show 3. ???shown 3.	sentence 24. sentences 10. indicator 9. indicators 9. paragraphs 6. paragraph 10. connections 6. equivalents 4. content 3. contents 5. text 21. texts 3.	table 9. items 12. type 7. position 6. important 5. samples 5. analysis 4. dependent 4. determined 4. example 4. indicating 4. predicate 4. ratios 4. topic 4.

catalan 30. acoustic 15. phonological 12. spanish-dominant 10. spanish 9. minimal 8. catalan-dominant 6. lexical 5. auditory 4. linguistic 4. sensitive 4. bilingual 3. phonemic 3. barcelona 4. based 4. . memorized 3. non-linguistic 3. priming 3.	repetition 19. effect 16. subjects 13. contrast 10 identification 8. recognition 8. phoneme 7. contrasts 6. discrimination 6. language 6. pseudo-words 6. memory 5. speakers 5. speech 4. ???appeared 3. distinguish 3. match 3. non-words 3. received 3. recorded 3. repeated 3. response 3. responses 3	words 33. word 15. pair 12. pairs 8. stimuli 6. stimulus 6. analysis 6. analyses 5. representation 5. representations 12. lists 9. list 4. experiments 3. experiment 4. lscp 3.	first 8. second 8. significant 8. tasks 8. vs 7. example 6. items 6. type 6. detailed 5. main 5. years 5. age 4. case 4. categories 4. direct 4. member 4. occurrence 4. participants 4. people 4. presented 4. results 4. study 4. categorization 3. classification 3. category 3. common 3. comparison 3.
normative 40. explanatory 32. skeptical 28. internal 27. behavioral 21. .mental 6. future 16. causal 15. rule-following 14. collective 10. observable 6. brute 7. exemplary 7. external 6.	behavior 39. condition 26. content 23. regularities 19. solution 17. reproduction 12. transcendence 12. members 10. norms 10. performance 10. critique 9. explanation 9. paradox 9. practices 9. argument 8. baseline 8. acquisition 6. constraints 8. collectivism 8. language 8. reference 7.	judgment 52. agreements 37. agreement 32. fact 38. facts 24. rule 23. rules 9. individual 18. individuals 16. state 14. states 31. form 13. forms 12. response 14. responses 11. wrpl 24. kl 8.	account 23. problem 23. life 22. community 19 methodological 15. person 14. appropriate 13. situation 13. words 13. activity 12. notion 12. following 11. group 11. particular 11. strong 11. apparent 10. function 10. important 10. indeed 10. point 10. question 10. similar 10. basis 9.

	???explain 8. present 8. provide 8. think 8. act 7. answer 7. seem 7. seems 7. argued 6. agree 6. appears 5. believe 5.		contents 9. instances 9. matter 9. claim 8. features 8. purely 8. terms 8. case 7. embodied 7. information 7. instance 7. learning 7. role 7. observed 7. reproduced 7. see 7.
normative 40. explanatory 32. skeptical 28. internal 27. behavioral 21. .mental 6. future 16. causal 15. rule-following 14. collective 10. observable 6. brute 7. exemplary 7. external 6.	behavior 39. condition 26. content 23. regularities 19. solution 17. reproduction 12. transcendence 12. members 10. norms 10. performance 10. critique 9. explanation 9. paradox 9. practices 9. argument 8. baseline 8. acquisition 6. constraints 8. collectivism 8. language 8. reference 7. ???explain 8. present 8. provide 8. think 8. act 7. answer 7. seem 7. seems 7. argued 6. agree 6. appears 5.	judgment 52. agreements 37. agreement 32. fact 38. facts 24. rule 23. rules 9. individual 18. individuals 16. state 14. states 31. form 13. forms 12. response 14. responses 11. wrpl 24. kl 8.	account 23. problem 23. life 22. community 19 methodological 15. person 14. appropriate 13. situation 13. words 13. activity 12. notion 12. following 11. group 11. particular 11. strong 11. apparent 10. function 10. important 10. indeed 10. point 10. question 10. similar 10. basis 9. contents 9. instances 9. matter 9. claim 8. features 8. purely 8. terms 8. case 7. embodied 7. information 7. instance 7. learning 7.

	believe 5.		role 7. observed 7. reproduced 7. see 7.
higher 130. european 65. east 62. communist 24. western 24. academic 19. eastern 19. socialist 18. social 17. political 16. state-socialist 14. recent 13. foreign 10. national 10. public 10. current 9. international 9. past 9. russian 9. established 8. future 8. private 8. west 8. cultural 7. difficult 7. senior 7. bureaucratic 6. conservative 6. economic 6. soviet 6.	education 123. countries 46. knowledge 27. agencies 24. institutions 24. europe 22. students 17. quality 14. report 14. support 14. practices 13. region 12. research 12. funding 11. role 11. science 11. studies 11. history 10. legitimacy 10. needs 10. status 10. teaching 10. academics 9. place 9. study 9. traditions 9. excellence 8. management 7. regimes 7. lack 9 disciplines 13. degrees 12. learning 8. societies 7 goal 12. processes 8. programs 8. standards 8. colleagues 7 level 14. co-operation 8. cooperation 6. economies 6. government 6. ideas 6. introduction 6. partners 6.	universities 49. university 31. faculty 23. reform 37. reforms 22. states 7. state 16. program 6. projects 19. project 6. systems 16. system 7.	significant 19. related 18. different 16. number 15. problem 12. similar 12. years 12. growing 11. involved 11. main 11. time 11. way 11 conditions 10. example 10. far 10. important 10. possible 10. largely 9. particularly 9. various 9. world 9. degree 8. extent 8. full 8. highly 8. issue 8. obvious 8. part 8. differences 7. elements 7. fact 7. former 7. limited 7. local 7. particular 7. people 7. position 7. post 7. problems 7. . terms 7. variety 7. actual 6. assurance 6. available 6. element 6. field 6.

	individuals 8. initiatives 8. society 6 student 6.. return 8. remains 7. change 6. means 6. becomes 8. changes 8.		latter 6. expected 6. massive 6. positions 6. principles 6. production 6. relatively 6. result 6.

The classification of the repetition types corroborates the finding that these four groups build the structure of a research article. In addition to the four types, it seems as if a fifth one can be distinguished, and this is repeating verbs. The following concordance line shows the repetition of the verb *understand*:

1.
*to a man's ability to **understand** chinese*
*argument that the system must **understand** chinese*
*the internalized systems example doesn't **understand** chinese in the sense*
*you know that other people **understand** chinese or anything else*
*but the man certainly doesn't **understand** chinese and neither do the*
*while a person doesn't **understand** chinese somehow the conjunction of*
*case that i fail to **understand** chinese*
*or computers none of which **understand** chinese the program notwithstanding*
*and bits of paper might **understand** chinese*
*manipulation system really does **understand** chinese*

*english and i do not **understand** the chinese and granted therefore*

*literal sense in which i **understand** english*
*o suppose that when i **understand** english i am operating with*
*that i am able to **understand** english and have other for*
*sense in which i **understand** english*
*the sense in which i **understand** english the issue would not*
*it the case that i **understand** english and a corresponding*

*my original example i **understand** the english and i do hardware and thus we can*

2.
*system and the system does **understand** the story*
*can literally be said to **understand** the story and provide the*

*in the room does not **understand** the story the fact is*
*explains the human ability to **understand** the story and answer questions*

3.
*simulate the human ability to **understand** stories*
*the sense in which i **understand** stories in english because a*
*a lesser degree i can **understand** stories in french to*
*need for this argument i **understand** stories in english to a*
*stories in english which i **understand** and they then ask me*

4.
*machine on the other hand **understand** nothing they are not in*
*you like but i still **understand** nothing*
*of a program and i **understand** nothing in the english case*
*in the case where i **understand** nothing*

***understand** the mind without doing neurophysiology*
JS

The verbs connect with two types of direct objects at a first level– *understand Chinese and English*. There is a second level of understanding discussed in the text connected with understanding a specific story and stories in different languages. Finally, cases of lack of understanding are picked up by the repetitions of the verb. It is fairly obvious, therefore, that the repetitions outline the way in which the problem is tackled– understanding of languages through understanding a specific story in two languages. This makes the repetition similar to the type called here **illustrative chains**.

The examples below show the repetitions of another verb–*seems.* A close inspection shows a completely different pattern from the one discussed above:

*1 could not have acquired any knowledge." Correspondingly, in our present case, it **seems** that knowledge of a languages grammar — can be acquired only by an organis*
*2 l be demonstrated when they come to be extensively applied. For the present, it **seems** that most complex organisms have highly specific forms of sensory and perc*
*3 atever properties are associated with the conditions of its use. Once again, it **seems** that language should be, for this reason, a most illuminating probe with w*
*4 lassical discussions of learning. As I have now emphasised several times, there **seems** to be little useful analogy between the theory of grammar that a person ha*
*5 he matter. Putnam's paper deals more directly with the points at issue, but it **seems** to me that his arguments are also inconclusive, because of certain incorre*

6 falsity, not a methodological one of states of investigation. To summarise, it ***seems*** *to me that neither Goodman nor Putnam offers a serious counterargument to*
7 most abstract sort. Speculating about the future development of the subject, it ***seems*** *to me not unlikely, for the reasons I have mentioned, that learning theory*
8 ered by Robert Ardrey and Joseph Alsop and popularised as a prophet of doom. It ***seems*** *to me that Lorenz's views on human aggression have been extended to near a*
9 ed not delay the study of the topics that are now open to investigation, and it ***seems*** *futile to speculate about matters so remote from present understanding. I*
10 acquisition of knowledge that gives due place to intrinsic mental activity. It ***seems*** *to me, then, that the study of language should occupy a central place in g*
11 lated in new ways and seen in a new light. For the first time in many years, it ***seems*** *to me, there is some real opportunity for substantial progress in the stud*
12 cesses that would have offended the common sense of earlier generations. But it ***seems*** *clear that this issue need not delay the study of the topics that are now*

While the repetitions of *understand* present the theoretical framework within which answers to the research question are sought, in this case the repeated verb serves the author to hedge his statements. In this sense it resembles the use of general research vocabulary. Further research into verbal repetitions confirms the fact that repetitions of verbs do not form a uniform group. Instead, they follow the pattern of one of the other repetition types. Therefore, they are better interpreted under the respective nominative categories.

Part Three: What does each chain of repetitions look like?

Below each category will be presented separately and specifics will be sought within the reference types and conceptual content.

Illustrative Chains

We start the discussion of the types of chains with the second type– concepts contributing to the analytical procedures in the article. Table 2.2 presents these repetitions:

Table 2.2. The Illustrative chains in the research articles

Article 1	***Article 2***	***Article 3***	***Article 4***
transition 70.	areas 39.	agreements 37.	systems 34.
transitions 70.	area 25.	agreement 32.	system 32.
shift 30.	neuron 13.	state 14.	grammar 62.

shifts 24. utterances 39. utterance 29.	neurons 39.	states 31. form 13. forms 12.	grammars 16. structures 20. structure 44.
Article 5	***Article 6***	***Article 7***	***Article 8***
indicator 9. indicators 9. paragraph 10. paragraphs 6. sentence 24. sentences 10.	words 33. word 15. pairs 8. pair 12. contrast 10. contrasts 6. stimulus ?? stimuli	universities 49. university 31.	symbols 38. symbol 13. system 28. systems 20. machines 20. machine 32.

As can be seen, the chains always include the plural and the singular form. The words illustrate experiments conducted by the researcher. The identifying uses are firstly introduced with an indefinite phrase, which is shortened or lengthened with various adjectives or post-positive phrases to name different referents. Finally, the chain flows into the plural one, because this is how the procedure can be described best:

*1. unequivocally that when a spoken **stimulus** is processed some of its*
*2. thus for example the **stimulus** s _ ba was followed*
*3. down the list by the **stimulus** seba these two words differ*
*4. headphones every seconds for each **stimulus** the participant had to decide*
*5. after the offset of the **stimulus** response time was measured from*
*6. from the onset of the **stimulus***
*7. created four counterbalanced lists of **stimuli** in the following way in*
*8. vs non-words and list of **stimuli***
*9. of the instructions played the **stimuli** off the hard disk and*
*10. method material all the **stimuli** used in the experiment were*
***11. stimuli** were presented through headphones every*
*12. a catalan speaker reading those **stimuli** at a pace of one*

The referents are different but the repetitions evoke the same type of object, in this case – stimuli for an auditory experiment. They are presented as different representatives (2 and 3), as a whole group (7,8,9,10,11,12), as an iterative representative (4), a specified member (1) and restructured objects (5 and 6).

The referential types may also include identifying uses, generic names, specified representatives or groups and substitutes. Each type enters the propositional function performed with the reference. Below we see identified uses which give concrete examples:

*kripke's **agreements** in judgment can be enacted*
*life and their constituent **agreements** in judgment*
*any account of how these **agreements** in judgment can be reproduced*

Next is a case when a general statement is made about the class and therefore, the complex is generic:

*the skeptical solution's insistence that **agreements** in judgment are explanatory primitives*

In the next set of concordances each use specifies and thus directs to subgroups of the noun class:

*words a set of relevant **agreements** in judgment proper to a*
*certain behavioral regularities manifest certain **agreements** in judgment*
*the appropriate behavioral regularities or **agreements** in judgment characteristic*
*the collective fact of the **agreements** in judgment characteristic of given*

The substitutes refer back to quoted examples:

*such **agreements** are not a matter of*
*of life and their constituent **agreements** in judgment represents a part*
*in fact come to the **agreements** in response that are supposed*
*life i e the behavioral **agreements** in terms of which such*
*forms of life and the **agreements** in judgment of which they*

Some repetitions classify in predicative uses, that is, present the objects as part of the class named with the noun:

*because they are normative **agreements** in judgment must transcend any*
*behavioral regularities that are **agreements** in judgment stand as explanatory*

From a different article, here is a case of what I call “iterative” reference:

*assigned two scores for each **transition** whether it was appropriately positioned*
*it lists the type of **transition** to each clause the semantic*
LR

Here the noun names several occurrences of the referent. The referential identity is different but the representatives belong to the same class named by the respective noun.

The process which we called “restructuring” is typical of this type chain. We see the concept restructured to name a range of aspects:

*1. that with the acceptance of **agreements** in judgment or behavioral regularities*
*2. consistent with his description of **agreements** in judgment but even i*
*3. provide one account of how **agreements** in judgment could be*
*4. skeptical solution's appeal to **agreements** in judgment does have causal*

The author is still talking of agreements in judgement, but 1 names a slightly modified notion–acceptance of such agreements, 2–description etc. This is accomplished by including the repetition into prepositional phrases.

As for shifts of the intension, the conceptual load can be characterised by proximity or overlap:

*the language **areas'** of broca and wernicke may*
*mostly outside the perisylvian language **areas'** cf*
*the neocortex and lower brain **areas***

Both Broca’s and Wernicke’s areas are lower areas in the brain, they are part of the perisylvian region and are related to language. For various reasons, different names have been given to each, but the objects in reality partially coincide. What unites them is their belonging to the concept “language areas in the brain”, as signified by the repetition. This is a case where the referential function clashes with the attributive one–one area is given different names for the purposes of analysing different functions of the component parts.

In the case of generic nouns we see a shift between the citation form and a representative of the class:

*1. the notion of a cohesive **transition** can make other contributions to*
*2. the incorporation of a cohesive **transition** into our centering account gives*

While 1 names the term as such, 2 picks up a generic name, which should be seen as a pragmatic shift.

Therefore, the illustrative chain tends to contain a great variety of referential types and a lot of restructuring, because it serves the operative purpose of naming the instrument through which the author seeks to find solution to the problem posed with the research.

Analytical/ Conceptual Chains

For this type of chain, we can see that active restructuring of the concept takes place through the repetitions. Most of the occurrences name a different aspect of the object under investigation. One of the articles contains this chain of repetitions:

*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*
*the goal of such **reform** ideas was to restore the*
*romelaer describes much better the **reform** processes in east european higher*
*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*
*europe the initial requests to **reform** higher education systems were*

The restructuring of the concept makes it obvious that the article is about the reforms in East European Universities. Generic reference and active restructuring are the major features of such chains. The objects named in this way are intensional–they do not pick out concrete objects but ideas and notions. Without being the same thing, the objects named with the repetitions belong to the same conceptual field: *reform efforts* are part of the *reform process*, obviously conducted by the *reform agents* and described in *reform documents*, following *requests to reform* the system. That is why the repetition serves the purpose of discussing an issue – the reform–in a number of aspects.

Concerning the form of the naming complexes in which the repetitions are included, they contain a number of nouns each of which restructure the referential set of the noun modified. The multitude of nouns combined in a phrase includes intensional content from all the participating lexemes. This is instrumental in developing new concepts where intensions are borrowed from a number of lexemes.

The complexity of the naming phrases also allows repetitions from different chains to be included in one phrase. For example, in the text we explored at the beginning of the chapter we established as the main chains repetitions of the *human*, *language*, *acquisition* and *device*. In this complex we see all of them:

of a specifically ***human language acquisition device***

Such concentrations of repetitions of high frequency reveal the direction of the argument. Sometimes parts of the complex are reiterated to discuss aspects of the concept; at other times the whole complex is picked up to remind of the direction of the argument:

explain how normative modes of ***behavior*** *can be assimilated and reproduced*
explanation of how modes of ***behavior*** *can be assimilated by those*

Repetitions of the whole complex are less frequent than the repetitions of the component parts. The phrases in their entirety also tend to occur at the beginning and the end of the text.

As for the referential identity of the phrases, the objects named are of the intensional type, i.e. unrelated to a thing in reality. When the phrase is repeated as a complex, the referential identity is ensured. Not so for the repetitions of parts of the complex. It is often the case that the naming complex is enriched to restructure the concept, as can be seen from the following example:

lexical parallelism *that is the repetition of*
lexical parallelism ratio *is t e determinable*
lexical parallelism indicator *given sentence sj through*

The basic concept is *lexical parallelism*, but the repetitions further evoke the derivative concepts "the ratio of lexical parallelism" and "an indicator of lexical parallelism".

In the following concordance five specimens of the class "brain" are named: the generic concept of brain, the brain of a Chinese person, an artificial brain called Ed and a brain simulator. The differences among these concepts are in the core of the description.

is often expressed that ***the brain*** *is a digital computer*

*at the synapses of **the brain of a native chinese speaker***
analogue of the human brain
*in consciousness and **the brain ed***
*the problem with **the brain simulator** is that it is*
IC

Intensions change to name different notions, as was shown in the sample text in isolated cases–one single use of *study* refers to a piece of research, while all the others refer to the process of investigating. The type of intensional change which furthers understanding of a concept is found in this concordance:

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

Unlike different meanings, in this case the word is associated with different pragmatic concepts: 1 features knowledge in university courses; 2-knowledge in general; 3–ideological knowledge; 4-an acquired bulk of knowledge; 5–competences developed at University; 6–enlightenment; 7-the personal range of what is known. The repetitions, therefore, serve to develop a variety of aspects associated with the concept in discussion.

That is to say that the analytical chains are also full of restructuring, as was the case with the illustrative chains, but unlike them they serve the purpose of enriching the intension. They also contain a great deal fewer identifying uses than the illustrative chains, because they do not pick up examples or specific cases.

Attributive Chains

The attributive chains contain mainly adjectives or attributively used nouns which maintain reference to a selected field. Below is an example of an attributive chain including the repetition of an attributively used noun:

SPEECH
*1) syllabic context presumably because **speech** perception by a phonetic module'*
*2) tract in the sense that **speech** perception implies recovering the vocal*

*3) hand liberman mattingly argue that **speech** perception can only function through*
*4) speech production but also during **speech** perception celsis et al*
*5) compatible with fodor's claim that **speech** perception is carried out*

*6) is activated not only during **speech** production but also during speech*
*7) deficits of **speech** production and **speech** perception respectively*

*8) interactive processing between levels of **speech** perception altman gernsbacher marslen-wilson tyler*
*9) to the motor theory of **speech** perception phonemic decoding is based*

*10) concerning **speech** production data on performance errors*
*11) semantic and phonological processes of **speech** production are both serial and*
*12) exclusively lead to deficits of **speech** production and speech perception respectivel*
*13) the vocal tract gestures' of **speech** production*

*14) approaches that may help explain **speech** emergence within a more general*
*15) **speech** emergence and perceptuomotor evolution as*
*16) kinds of preadaptations related to **speech** emergence might still be found*
*17) autonomy and phylogenetic discontinuity of **speech** emergence only indirectly*
*18) since straightforward evidence on **speech** emergence in hominids is unavailable*

*19) can also differentiate between **speech** stimuli on the basis of*
*20) thus neonates can discriminate **speech** stimuli from nonverbal auditory stimuli*
*21) there are indications that **speech** stimuli are received by the*
*22) betz auditory evoked potentials for **speech** stimuli in neonates predict*

*23) level not directly observable in **speech** behavior*
*24) describe finite samples of **speech** behavior was related to his*

These items do not pick out specific objects or concepts but a set of features. Therefore, they do not "refer" in the proper sense of the word, but direct to a field. Unlike the case with *reform* above, it is not the case that speech as such is being discussed in its different aspects. Rather the inventory used for the discussion of the central concept is within the domain of speech–its perception and production, the stimuli and behaviour are involved in the study.

The attributes occur in phrases with different nouns and thus create lists of all the objects from a thematic field included in the discussion:

MENTAL

*genuine understanding and other **mental states** the first thing to*
*isn t confined to simulating **mental operations** by any means*
*a precise well defined thesis **mental processes** are computational processes over*
*philosophy and our **mental life** in mind language and*
*to reproduce and thereby explain **mental phenomena** and i believe*
*specifically mental about **the mental***
NSK

In the end, the adjective is nominalised in the definite phrase to highlight the significance of the thematic area.

Other adjectives are found in fixed phrases with a noun, which obtains such a set value that it is further restructured, as can be seen from the examples below:

LEFT
*is a consensus that the **left** hemisphere is mainly involved*
*marcotte morere report **left** hemisphere dominance in only*
*suggests a causal relation to **left** hemisphere language dominance found in*
NCK

The attributes can also be restructured by adverbs:

*those systems that are **genuinely mental** from those that are not*
*of what is **specifically mental** about the mental*
IC

As for intensional shifts, when the adjectives link up with different nouns, some of the modifications in the resulting complexes feel like pragmatic shifts:

*same bilingual education as the **catalan natives***
*then we created sixty-four **catalan pseudo-words** yielding thirty-two minimal pairs*
LR

In the first line the nationality adjective is used, while in the second the reference is to the language.

Therefore, the attributive chains do not “refer” to objects, but direct the attention to intensional fields characterised by the features of the respective adjectival or nominative class.

General Research Vocabulary

This group includes nouns, adjectives, numerals, adverbs, verbs and other morphological classes. The semantic nature of most of these words presupposes membership in the group "research vocabulary". Some of the examples include *problem*, *table*, *type*, *position*, *goal*, *sample*, *role*, *status*, *case*, *sense*, *organisation*, *fact*, *example*. The latter two also form part of the phrases *in fact* and *for example.* Some of these lexical items are only associated with a singular form, others–with plural. From the *pluralia tantum* group a few examples are: *data*, *practices*, *conditions*, *properties*.

The type of repetition called "general research vocabulary" also includes contractions, such as *cf*, *ff*, *ibid*. and others. The reference of these items is very flexible–they take meaning from the phrases they appear in. In the case of the contractions, it is a matter of sticking to a convention rather than a question of naming. Their function is to explain the research procedures to which the material is subjected.

Some lexical items belong to more than one morphological class. Examples are *account*–noun and verb, and *lack*–noun and verb.

Many texts include the participle *related* as the articles often need to discuss relations between concepts and events. Likewise, adjectives such as *similar, identical* and *dependent* perform similar relational functions. Another semantic group, with an evaluating character this time, includes the adjectives *significant* and *important*. Temporal features are bestowed by repetitions of *recent* and *current*. Emphatic functions are performed by the adjectives *specific, certain* and *far,* and the adverbs *particularly, largely* and *rather*.

A comparison between general research vocabulary chains and analytical chains sets off significant differences between the two. Firstly, explore an analytical chain:

be taken as a ***brute fact*** *wrpl*
are a matter of ***brute fact***
by ***brute fact*** *i understand kripke to mean*
to explain beyond the ***brute fact*** *of agreement of responses that*
collective fact *as solution after concluding that*
found in the ***collective fact*** *of the agreements in judgment*
to individuals to a ***collective fact*** *that is observed as*
same situation to the ***collective fact*** *which is that members of*

*not simply describe the **individual fact** of jones's supposed conformity t*
*get us from the **individual fact** that jones is behaving in*
*be found in the **individual fact** of those states of the*
*still is no **internal fact** of the matter to consider*
*apparently is no non-regressible **internal fact** about the purported rule-follower*

The noun is repeated in phrases with four attributes: *collective*, *individual*, *internal* and *brute*. The complexes are obviously fixed phrases and function as terms in the respective area, which is confirmed by the fact that the phrases recur in their entirety. Even within this limited concordance we notice two other repeated items–*agreement* and *judgement*. Quite unlike these repetitions are the ones in the general research vocabulary chain from another article:

*come to grips with the **fact** that all normal humans*
*of language lies in the **fact** that in this study it*

*an otherwise intelligent ape a **fact** that was emphasised quite correctly*
*we take note of the **fact** that this militant anti-psychologism is*
*component is explained by the **fact** that all the natural measures*
*is nothing surprising in the **fact** that languages contain rules of*
*from the fact **that** a phrase structure*
*of his materials beyond the **fact** that the savage mind attempts*
*enough to account for the **fact** that a specific grammar is*
*phonemic systems but in the **fact** that a fairly small number*
*the **fact** that the mind is a*
*the **fact** is that this concept has*
*of linguistic structure but the **fact** is that we have for*
*recognition of this **fact** though formulated in entirely*

*man's nature is attributable to **fact** and logic and to*

No fixed phrases can be found in this concordance and no structural relations between such complexes. You will also have noticed that with one single exception–in the last line–the reference of all the cases is of the type we called "substitute".

The prevailing referential type is specific-ness, and quite frequently–substitute. As for restructuring, it would seem as if the very lexemes lend themselves to restructuring other lexical items rather than to being restructured themselves. The following example reveals the nature of qualitative restructuring. The repeated word is *properties* and it restructures several nouns in turn:

PROPERTIES
*close relation between innate **properties** of the mind and features*
*the moment about the general **properties** of the underlying phrase structure*
*empirical question whether the **properties** of the language faculty are*
*way but it is the **properties** of the systems of rules*
*do share many of the **properties** of human gestural systems and*
*the essential and distinctive **properties** of human intelligence than through*
*then argue that these **properties** of natural language are acquired*
*aims to elaborate the formal **properties** of any possible human language*
*succeeding generation along with whatever **properties** are associated with the conditions*
*specifies correctly or incorrectly certain **properties** of bad languages*
*and animal language are the **properties** of being purposive syntactic and*
*attempt to explore the innate **properties** that determine how knowledge is*
*assumes that the only significant **properties** of language are that*
*to show that the specific **properties** of grammar say the*
*can be obtained concerning the **properties** of grammars and conclusions regarding*
*and surface structure the specific **properties** of grammatical transformations the*
NC

Other repetitions restructure quantitatively. No semantic links are visible from this concordance, but the linking power of this type of chain lies in the fact that quantification is important for the argument.

NUMBER
*a **number** of other types of semantic*
*condition but also by any **number** of other semantic relationships is-a*
*table we see a high **number** of complete shifts*
*estimation of a possible maximal **number** of complete shifts we would*
*the **number** of complete shifts is comfortably*
*question raised by the high **number** of nonexplicit cbs found in*
*model because of the high **number** of nonexplicit cbs in this*
*is diluted by the high **number** of nouns in ui in*
*in this case the average **number** of nouns per clause in*
*containing inferable cbs and the **number** of null transitions*
*to attempt to minimize the **number** of null cbs reasoning that*
*is the particularly high **number** of transitions to utterances containing*
*no cb we hand-tabulated the **number** of transitions of each type*
*computational linguistics volume **number** c _ association for computational*
*is this **number** an accurate estimation of the*
*cohesive transition there are a **number** of aspects of the proposal*
*they test a **number** of congurations of parameters*
*this corpus contains a high **number** of discourse elements that are*
*with shifts in cohesion this **number** at least gives us a*

*completed the group exchanged a **number** of opinions and pieces of*
*ui table reports the total **number** and percentage*
*ishara and nagao induced a **number** of possible relations between bridging*
*continue followed by a respectable **number** of retains and a very*
*over messages then the total **number** of such scores was scores*
*confounding factors that make this **number** higher than is actually appropriate*
*more problematic is that a **number** of these missing words e*
*the **number** of times a keyword appears*
*a **number** of useful related strategies that*
*have been elaborated in a **number** of works nariyama nakaiwa and*

The intension of the repeated words is shaped by the context of each recurrence. Observe, for example, the following chain, where each "first" belongs to a different ordinal scheme:

FIRST
*average reaction times on the **first** and second occurences of words*
*of the relationship between the **first** and the second member of*
*in the **first** batch of analysis we excluded*
***first** for the spanish group the*
*the **first** items of the lists were*
*was the same token the **first** or the second member of*
*in barcelona northern spain the **first** population consisted of people raised*
*the direct access hypothesis was **first** proposed to our knowledge by*
JS

The repeated nouns are used with different articles, as the concordance below shows, and this fact sets off a different meaning of the lexeme. The first group reveal the meaning "senses", the second–"meaning", the third one–"according to a particular interpretation",the fourth–"to a certain degree" and the sixth one includes the noun in an idiomatic phrase.

SENSE
(1)
grammar from the data of sense
consistent with the data of sense
also differentiate the data of sense into those utterances that give
(2)
is also syntactic in the sense just defined as in
corresponding to competence in the sense in which competence is
is descriptively adequate in the sense described in lecture can
in this sense then both human language and
in this sense the idea of a triangle
(3)

conclusions is correct in his sense of experimental test namely
matter in motion in his sense
dualist in the cartesian sense he argued not very persuasively
a language in the humboldtian sense namely as a recursively generated
a precondition in the kantian sense for linguistic experience and it
of intellectual development in piaget's sense but rather slow progres s

(4)
there is an obvious sense in which any aspect of
view that in some sense the mature mind contains ideas
behaviour in a sufficiently loose sense
language evolved in any strict sense from simpler systems but he

(5)
would have offended the common sense of earlier generations
our in the most general sense of purposive walking is also
language in the true sense and not the mammals thorpe
dimension insofar as it makes sense to speak of continuity in
NC

This type of use is only possible because of the semantic flexibility of the items in the group of general research vocabulary. The repetitions are quite distant from each other and my guess is that they do not feel as repetitions for the reader, i.e. the reader would not think that the author is repeating the same lexical item.

In conclusion, the general research vocabulary chains allow researchers to reveal the details of their methodology. The referential types are concretely identified or specified entities, with quite a few substitutes which help refer back to complexes. The changeability of the intensions is a function of the semantic nature of the items which occur in this type of chain. This is one reason why good authors do not seem to avoid such repetitions.

Part Four: Conclusions about the repetitions in research articles

In our effort to establish the role of repetition in building coherent texts we set out to find answers to four questions. Based on our analysis of the corpus of the research articles, we answer these questions:

1. How significant is the repetition of word forms in each separate genre?
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 intension and the form?
4. Does the nature of repetition change with each genre?

How significant is the repetition of word forms in each separate genre?

We evolved a procedure for normalising the number of repetitions by presenting them per one thousand words. This index for the texts in our corpus is, as follows:

Table 2.3. Index of the repetitions in research articles

	LR	PH	NC	HE	IC	JS	NCK	IFC	
Word forms	1333	2400	8175	10 253	10420	11461	14563	18978	77583
Different	411	738	1713	2510	1649	2182	2799	4316	
Repetitions	922	1662	6462	7743	8771	9279	11764	14662	
Normalised index	691	692	790	755	841	809	807	772	

As can be seen, more than half the words are repeated in each text and more than 30% of the full lexical words. Such a scope cannot but be significant for the texts and if authors refrain from repeating, no evidence of this is to be found in our corpus.

Is it the case that repeating word forms creates text structures?

Four types of repetition chains were established in research articles, each characterised by its specific form, referential type, restructuring and intensional composition.

Table 2.4. Types of repetition chains in the research articles

Attributive chains	Adjectives or attributively used nouns are repeated. Recur in complex phrases and co-refer to terms; or recur with a thematic list of nouns.

Conceptual/analytical chains	Nouns or verbs are repeated. The referential types differ to single out identified representatives, to generalise with generic names, classify or specify. The restructuring is active and reveals a range of aspects of the concept under review.
Illustrative chains	Nouns are repeated. Two parallel chains of plural and singular forms are found in a variety of referential types. They reveal the concepts which are instrumental in conducting experiments, performing mental surmises etc. in order to reach conclusions about the issue under investigation.
General Research vocabulary	Nouns, verbs, adjectives or abbreviations are repeated. The prevalent referential type is specificity, quite typical are substitutes. The intensions are flexible and project a number of meanings, but they do not feel like the type of repetition which needs to be avoided.

The naming is due to complex phrases including intensions from a number of lexemes. The restructuring helps include various aspects and the repetitions make the links among the parts of the naming complexes lucid and easy to understand. The different referential types allow performing various propositional functions – concluding, exemplifying, identifying etc.

It is also fairly obvious that the taxonomy of the repetition chains reflects brilliantly the structure of the research article – a concept is investigated through an analytical framework, using various methods; multiple theoretical fields are outlined. Thus the repetition types indeed form structures which shape the skeleton of the genre.

What is the role of each side of the naming process: the referent, the intension and the form?

The intensions of the recurring items appear in development through the text, which seems to be the major purpose of a research article – to uncover a concept in its rich diversity. The concept is enriched in several different ways. Restructuring adds new components, while pragmatic shifts include other aspects of the problem. As the intension is representative of the knowledge of the author of the concept, each author makes sure they present as much as possible, within the boundaries of the study.

The intensional shifts which have been observed can be classified under the following headings:

1. A different bulk of the same concept
2. Concepts differently defined
3. Different aspects of the same concept
4. Specific groups within the same group
5. A group concept
6. The citation form of the term

Most of the terms named in the research articles do not link up with actual referents in reality. That is why they are sometimes called "intensional names". If we assume that referents are non-existent and the intensions are changing, then what indicates the identity of the named concept? In terms of our study, it has to be the lexical form. That is why repetitions are so numerous and affect the major concepts in the article. Undoubtedly, we may also introduce the idea of an intangible referent, which does not exist in reality and several studies speak of such referents.

The restructuring allows an in-depth discussion of the objects and concepts. The basis for it, needless to say are the repetitions.

The different referential types, as was shown are instrumental in performing different propositional functions. Authors tend to use repetitions as a point of orientation for new representations.

Having reviewed the repetitions in research articles written by established authors, for the purposes of comparison, we now take a look at work written by students. Eight articles are selected from unedited proposals for publication in student journals.

Part Five: How Do Inexperienced Authors Handle Repetitions?

Firstly, the comparison between the index of repetitions in the students' and established the authors' writing shows that the inexperienced authors repeat 30% less often than the experienced academics. In fact, the upper limit of repetitions in the students' articles barely reaches the lower limit of the researchers'.

Table 2.5. Repetition indexes in inexperienced writing

	Experienced scholars	***Students***
Number of repetitions	691- 841	475-632
	Research articles	

Secondly, in the students' articles it is hard to distinguish types of chains. All the chains repeat the same word or phrase in the same environment, without restructuring, intensional enrichment or different referential type, as can be seen from the table below:

Table 2.6. Repetition chains in inexperienced writing

Illustrative	**Analytical**	**Attributive**	**General Research vocabulary**
Image / Images - recurring, stereotypical Jobs / job female, male, the daughter's Stereotype – my own, several, numerous	Occupation the most widely spread not so traditional Characters Male Female 3, most Countryside South-eastern Inhabitants The reality Stereotyping	British – life, novelist, team, television English – county, countryside, ladies Female – characters, jobs, space Horse – racing, breeding Midsommer Murders	Fact Far Most Appear obviously

Identity + Gen Issues Gaps No Identities Sequence of Transformed Lost	American-ness Subject of Thinking about Trip into Time – undefined, instants of , at the same? Night ? Century Call (v) to, I would, I call Art – forms, of painting, works Life – meditations on Greatness - of sea, of dream, of wrath Celebration – a real c of cultures, of c.	American – literature, cultural history, identity, writing Undefined – time, result, one	more
World – his, the new, the old Puritans - the p's sense of exceptionalism Settlers – puritan, protestant Invasions – west coast I, Strategy of I America – west coast of A, discovery of, conquest of A Identity – their, I construction, lost i. Eternity – kingdom of E	Myth – the myth of Eden, a powerful myth, the birth of a myth, reviving the myth, myth-dwellers Imagination – mythological, a string of I, Colunbus's I. Eden – myth of E, archetype of E.	American settlement & American colonization, A land, Coast – west, east New – world, continent, born child Mythological – imagination, subconscious impulses Pscychological – explanation, trandsformation, conception Religious – filter of r & other conceptions, R & mental concept	Way – their own way, no, this, non-human way

Students – 10th class students, communication with, interaction with, experience Teacher – the role of the teacher, & textbook, a more teacher centered class room Textbook – role, literature , English through literature t. Unit – part of a, English through Literature U	Communication – components, amount of two way c, this, c in all it aspects	Cultural – assumptions, teaching c. context, Hofstede's c. dimension, another dimension of c diversity, areas of c. diversity, areas of cultural variety English – through literature, tbk	

The repetitions are classified into the respective types according to the meanings sought by the students, not because the chains show the characteristics of each group. Even so, it appears that specific groups are altogether missing–more often than not, of general academic vocabulary. The effect of this on the writing is that the description of the analytical procedures taken to explore the topic in question remains underdeveloped.

Below is the analytical chain for one of the articles:

ETERNITY

*looking for the kingdom of **eternity** puritans approach the east coast*
*looking for the kingdom of **eternity** puritans approach the east coast*

*america as their kingdom of **eternity** where they can re-construct their touch with the kingdom of **eternity***
*america as their kingdom of **eternity** where they can re-construct their touch with the kingdom of **eternity***

As can be seen, the repetitions do not restructure or present different aspects of the problem, but directly repeat the same phrase over and over. Thus the informative value of the repetitions is significantly decreased. In general, the major type of repetition chain in most cases includes a phrase from the title repeated without a change in reference type, restructuring or naming aspects of the problem.

Only one of the students varied the reference of the phrases in which she included the repetitions of the word *identity*:

IDENTITY

***identity** is not a modern issue*

*modernism in american writing **identity** issues it is just a*
*back to the issue of **identity***

*bridge continental ideological and mainly **identity** gaps and turn the*

*name and as if no **identity***
*being american or not **identity** is a matter of choice*

*and yet so indisputably american **identity***

This indicates that the author creates bridges between the concept of identity in general and identity gaps-which is a case of restructuring. Then she denies identity with the phrase *no identity* and finally identifies American identity. This shows one sort of argumentative development and it makes it clear what the author is trying to say with the essay.

In one of the texts a temporal chain–of the noun *night*–is found, which makes the text sound fictional rather than research.

In general, classifying the repetitions into the respective type of chain helps determine the quality of the text. A missing chain of general substitutes shows that the methodology is not clarified, a lack of restructuring-that the concept is not viewed from different aspects; an insufficient number of specified uses would indicate that the text does not give enough examples; contrarily, missing generic name-inability to draw conclusions.

In conclusion, we can say that experienced authors repeat words more often than inexperienced ones. However, what matters is not only the quantitative measure but the way the repetitions are included in phrases.

Part Six: Summary Machine

The pattern of repetitions established above can help summarise research articles. Let us take randomly an article from a research journal and classify the repetitions according to the types described above.

The frequency list with the function words removed is as follows:

DOCUMENTS	*97*			*BECAUSE*	*8*
CORPUS	*52*			*COMPANY*	*8*
INDUSTRY	*52*			*AVAILABLE*	*7*
TOBACCO	*45*			*FINDINGS*	*7*
WORDS	*28*			*FIRST*	*7*
SAMPLE	*27*			*GENERAL*	*7*
DOCUMENT	*26*			*HAND*	*7*
REFERENCE	*20*			*HOW*	*7*
TIDS	*20*			*LONG*	*7*
CORE	*18*			*MOST*	*7*
EXTERNAL	*18*			*PARALLEL*	*7*
PERCENT	*18*			*RATES*	*7*
RHETORICAL	*17*			*SEE*	*7*
CORPORATE	*16*			*SHORT*	*7*
PUBLIC	*16*			*SMOKING*	*7*
SAMPLING	*16*			*TEXT*	*7*
USED	*16*			*TIME*	*7*
AUDIENCE	*15*			*TWO*	*7*
RESEARCH	*15*			*TYPES*	*7*
AUDIENCES	*14*			*USE*	*7*
DECADE		*14*		*ADDITIONAL*	*6*
EXPLORATORY	*14*			*BLILEY*	*6*
HEALTH	*14*				
INTERNAL	*14*				
LINGUISTIC	*12*				
TEXTS	*12*				
WORD	*12*				
ANALYSIS	*11*				
CROSS	*11*				
MARKETING	*11*				
NAMED	*11*				
SIGNIFICANTLY	*11*				
TERMS	*11*				
CASES	*10*				
FIFTY	*10*				
LANGUAGE	*10*				
WOULD	*10*				
BROWN	*9*				
CONTENT	*9*				
DRAFT	*9*				
MANIPULATION	*9*				
PRODUCT	*9*				
TABLE	*9*				
ASSOCIATED	*8*				

Classified into the four types of repetitions established above, this is what we get:

Table 2.7. Repetition chains for the Summary Machine

Analytical/conceptual	*Illustrative*	*Attributive*	*General Research vocabulary*
INDUSTRY 52	DOCUMENTS 97	CORE 18 REFERENCE 20	CORPUS52
MANIPULATION 9	DOCUMENT 26	EXTERNAL 18	TIDS 20
PRODUCT 9		TOBACCO 45	SAMPLE 27
	AUDIENCE 15	CORPORATE 16	PERCENT 18
	AUDIENCES 14	RHETORICAL 17	SAMPLING 16
	WORDS 28	PUBLIC 16	RESEARCH 15
	WORD 12	EXPLORATORY 14	DECADE 14
		HEALTH 14	ANALYSIS 11
	TEXTS 12	INTERNAL 14	SIGNIFICANTLY 11
	TEXT 7	LINGUISTIC 12	TERMS 11
		MARKETING 11	
		DRAFT 9	
		CONTENT 9	

The word which is repeated with great restructuring of the notion is *manipulation*:

*1 and to estimate the extent and prevalence of **manipulation**; and (3) analyze manipulation in order to classify it and develop means to identify similar mani*
*2 of all eligible TIDs, regardless of whether they were expected to contain any **manipulation**. Finally, we compiled a corpus of parallel texts of particular rhe*
*3 n ("deception") may have occurred and to estimate the extent and prevalence of **manipulation**; and (3) analyze manipulation in order to classify it and develop*
*4 would prove of heuristic value in identifying potential linguistic markers of **manipulation** and deception. However, about only 3.5 percent, or 28 documents ou*
*5 c features of this unique set of texts; (2) identify TIDs in which rhetorical **manipulation** ("deception") may have occurred and to estimate the extent and pre*
*6 rrence of linguistic characteristics of interest in the analysis of rhetorical **manipulation**. Because many of these characteristics might occur at low frequenc*
*7 from those text types we considered relevant to (i.e., subject to) rhetorical **manipulation**. This reference corpus is a stratified random sample of all eligib*
*8 yze manipulation in order to classify it and develop means to identify similar **manipulation** in other industry document sets. Of course, these objectives requi*
*9 erence corpus was to create a comparison set of TIDs from among those in which **manipulation** potentially occurred (but did not necessarily occur), from which*

So the author picks up for discussion the extent and prevalence of manipulation, potential linguistic markers of manipulation, analysis of rhetorical manipulation - all through restructuring the repeated noun prefaced by prepositional phrases. Additionally, we notice generic uses as well as uses negating the existence of manipulation (*any manipulation*). This type of repetition–actively restructured to explore a number of aspects of a concept-is typical of the analytical/conceptual chains, which leads to the conclusion that this is the notion explored with the research.

Another notion restructured in the course of the article is *product*:

*1 at a significantly lower rate in the 1990s, while the words carton, pack, and **product** are used at significantly higher rates in the 1990s. This pattern may in*
*2 both internal and external to the industry provided increasing proof that the **product** was a major menace to human health. Toward the end of the fifty years, t*
*3 the Brown Corpus, nor that the lexicon of TIDs merely reflected the names and **product** terms of the industry. The use of words in industry documents is consis*
*4 in the 1990s. This pattern may indicate less focus in the industry on what the **product** is and more emphasis on its packaging for sale. The cluster of disease*
*5 e tobacco industry was a well respected, corporate, global citizen marketing a **product** that was often associated with relaxation and hardiness. During the cou*
*6 e displayed in Table 5. The first cluster contains the words of the trade: the **product** and its components, words for the act of using it (smoke, smoking, smok*
*7 rent research on TIDs pertains to revelations concerning marketing strategies, **product** design, or deception in reporting scientific evidence (Bero 2003), the*
*8 roup consists of words with applications to research, both market research and **product** research, and thus represents a combination of the marketing and health*

9 oker), and company names. The second group shows the vocabulary of selling the ***product****, including brand names and marketing strategies (blend, flavor, lights,*

Obviously the author discusses product design, research and terms. The reference of the respective phrases ranges from citation (in example 1), to substitute and generic noun.

The other highly frequent term is *industry*, which occurs with the attribute *tobacco*.

The general research vocabulary includes *corpus* and *TIDS*, the latter– introduced as the contraction for "tobacco industry documents". From the attributive list it becomes obvious that there is a *reference* corpus and a *core* one. The illustrative chains reveal that *texts* from *documents* are analysed meant for different *audiences* in terms of the *words* used there.

Our analysis of the types of repetition chains showed that the research framework is revealed by the noun which occurs in parallel chains of plural and singular forms. In this case we establish that this is *documents*. The attributive chains orient us towards the tobacco industry. The conceptual chains reveal that the concept under investigation is *manipulation.* The research vocabulary tells us that corpora are used for the analysis. All of the above leads to the following summary:

Manipulation is explored in documents from the tobacco industry, using different types of corpora.

Compare our guesswork based on the classification of repetitions with the abstract of the paper:

As a result of litigation over the past decade, major tobacco companies were compelled to make public a broad range of previously confidential documents. We have created a series of corpora from the tobacco industry documents (TIDs) for three purposes: (1) to establish baseline descriptions of various linguistic features of this unique set of texts; (2) to identify TIDs in which rhetorical manipulation ("deception") may have occurred and to estimate the extent and prevalence of manipulation; (3) to analyze manipulation in order to classify it and develop means to identify similar manipulation in other industry document sets. Our three part corpus creation strategy employed rigorous sampling methods. First, we drew a limited sample from the largest collection of TIDs, to determine a representative classification of text types and to estimate their proportions within the overall body of texts. Then, we created a reference corpus (500,000+ words) constituting a stratified random sample of all TIDs, whether or not they exhibit manipulation. Finally, we compiled a corpus of texts presumed to exhibit rhetorical manipulation. We assumed that multiple drafts of a text or versions of a text prepared for different audiences constituted rhetorical manipulation. This article presents our experience with the sampling methods utilized in this corpus-building process and our findings regarding text types comprising the reference corpus.
Keywords: corpus linguistics; rhetorical manipulation; text sampling methods; tobacco control

Except for the details, the major features of the article were caught through our classification of the types of chain. Therefore, we can conclude that by simply establishing the structure of repetition chains summaries of articles can be generated with a degree of accuracy. Additionally, this fact corroborates our claim that lexical repetitions form the framework of a text.