LEXICAL COHESION IN POPULAR VS. THEORETICAL SCIENTIFIC TEXTS

Radek Vogel

Abstract

Lexical cohesion, i.e. selection of lexical items which are in some way related to other lexical items in a text, contributes importantly to creating the texture, as defined by Halliday and Hasan (1976), and increases the overall coherence of the text. The article looks into different devices of lexical cohesion employed in scientific texts, namely reiterations (including exact repetitions and synonyms), use of superordinate expressions and general terms, and marginally also collocations. It is built upon an assumption that two basic types of scientific text, popular and theoretical ones, display some differences in the ratios between these lexical cohesive devices as well as in the frequency of their occurrence in each individual type. The analysis has been undertaken on two corpora of texts dealing with various topics in physics, comparing a text which is highly theoretical with one covering the same respective topic, adapted rather for didactic purposes or purposes of popular science. Lexical chains reveal the cohesive links between thematic elements of the texts and highlight the differences between individual texts matched in pairs. Particularly the use of general and superordinate lexemes in contrast to exact repetitions has proved to be the feature distinguishing between the two text varieties.

Key words

antonymy, direct cohesive chains, generalisation, lexical cohesion, popular scientific texts, repetition, specification, superordinates, synonymy, theoretical scientific texts

1 Introduction

Authors of written texts, since they cannot rely on immediate interaction with recipients of messages incorporated in their texts (unlike the possibility to ask for clarification in most types of spoken discourse), must plan and construct the texts carefully to secure their comprehensibility. For this reason, they employ a repertory of lexical, grammatical and structural means to express the relations between text components clearly and communicate thus the content and purpose of the text to the readers successfully. As Dontcheva-Navratilova (2006: 51) puts it concisely, "The writer´s role in this interaction is to anticipate the reader´s reaction and to use different signals and strategies to guide him/her in lifting the range of possible interpretations and support the intended interpretation."

This paper looks into one type of means which achieve connectedness and ensure understanding of written texts, namely the devices of lexical cohesion. Moreover, since the distinction of the medium of communication does not provide a complete characteristic of the considerations that must be made while constructing a target text, the remaining two aspects of the Halliday´s triad of register variation

(1978) have to be defined as well. The tenor, i.e. mutual relationship between participants in the act of communication, is realised as a certain level of formality or informality. The present paper assumes that two types of scientific text identified by the specific levels of (in)formality and anticipated readership, viz. popular and theoretical ones, reveal different frequencies of occurrence of and ratios between types of cohesive devices, including lexical ones. To specify also the third aspect, the field or province, the present analysis has been carried out on two corpora of texts discussing various topics in physics, contrasting theoretical (or academic) texts with their rather popular (or didactic) counterparts dealing with the same topics.

2 Lexical cohesion: its function and devices

Discourse in any field of activity is characterised by existence of two relations, namely semantic connectedness, referred to as coherence, and syntactic connectedness, known as cohesion (Urbanová 2008: 83), which is also the focus of this paper. Contrary to coherence, dependent on the result of interplay between the chosen cohesive features and subjective factors associated with the personality of the recipient, cohesion is a more objective and identifiable concept. Halliday and Hasan (1976: 293) stress its fundamental importance for the construction of texts: "Typically, in any text, every sentence except the first exhibits some form of cohesion with a preceding sentence, usually with the one immediately preceding. In other words, every sentence contains at least one anaphoric tie connecting it with what has gone before."

Halliday and Hasan distinguish two basic types of cohesion, viz. lexical cohesion and grammatical cohesion (Urbanová 2008: 83-84, Halliday and Hasan 1976: 303). Expanding this division, some linguists also point out to structural cohesion as a specific type of cohesion (cf. Dontcheva-Navratilova 2005), although Halliday and Hasan acknowledged its role as well.

Lexical cohesion is thus one of the main types of cohesion. Halliday and Hasan distinguish two related aspects of lexical cohesion – **reiteration** and **collocation** (1976: 318). The former one, reiteration, is realised either as the repetition of a lexical item or the use of a synonym (Ibid.). Since reiteration occurs in the context of reference, "a reiterated lexical item is accompanied by a reference item, usually *the* or a demonstrative." (Ibid.: 318-319)

Urbanová (2008: 84-85) lists two lexical devices among cohesive devices, viz. **reiteration**, which is seen as a repetition of key words or derivatives thereof, and **sense relations**, especially synonymy and antonymy. Similarly, Dontcheva-Navratilova (2005) employs a classification of subtypes

of lexical cohesion which is derived from semantics, namely various sense relations. Urbanová´s conception splits reiteration, as defined by Halliday and Hasan (1976) into two, separating synonymy and joining it with other semantic relations into a larger set based on similarity of meaning. As Urbanová (2008: 84) points out, lexical cohesive devices contribute to approximation of meaning ("významové zpřesnění"), explanation and connectivity of the content of text parts, whereas the tools of grammatical cohesion help to achieve surveyability and comprehensibility via arrangement of the grammatical structure of a text.

Zmrzlá (2009: 39-40) attempted at a synthesis of the above-mentioned classifications of grammatical, lexical as well as structural cohesion by Halliday and Hasan 1976, Tanskanen 1995 and Dontcheva-Navratilova 2005 and combined them into a single hierarchy. Leaving grammatical and structural cohesion aside, lexical cohesion is divided into reiteration and collocation (identically with Halliday and Hasan 1976), and reiteration is further divided into repetition, equivalence, generalisation, specification, co-specification and contrast, while collocation has two subtypes, viz. ordered set and implication.

3 Theoretical and popular scientific discourse

3.1 Lexicon of exact science

Knittlová (1990: 26) suggests the division of the style of science and technology ("styl vědy a techniky") into two distinct branches, the theoretical and the practical (or didactic) one, and mentions also Mistrík´s classification of the so-called didactic ("naučný") style into the scientific and popular branches. The lexicon of the style of science is characterised (Ibid.: 27) by conceptuality, marked use of nouns and adjectives, use of terms with a clear and narrow denotation, absence of expressive lexemes, exclusiveness and therefore high repetitiveness of lexis, resulting in quite a stereotypical vocabulary, and semantic condensation with preference of noun groups (Ibid.: 40, 49). The popular scientific style differs from its theoretical counterpart by less specialised terminology, inclusion of colloquial expressions, higher occurrence of paraphrase, expressive and personality features, etc. (Ibid.: 28-29).
Knittlová (1990: 47) also mentions Galperin´s (1977) observation that there exist distinct styles of exact sciences on the one hand, and humanities on the other, each with their typical means though it is possible to identify a convergent trend.

Schmied (2006) reports on a corpus-based research dealing with complexity in lexis and syntax, which investigated two types of discourse: specialist discourse (on-line databases and peer-reviewed

publications) and non-specialist discourse (namely popular adaptations of academic articles in *New Scientist*) (2006:144). Together with an analysis of syntactic complexity also lexical complexity was in the focus of the research. The author hypothesised that non-specialist texts would have fewer unknown nouns and more general nouns, whereas the specialist texts would be more complex syntactically and lexically. The analysis was carried out with the use of a computer software and followed the methods of corpus linguistics. It proved the above-mentioned hypotheses and usefulness of the number of unknown words (according to Word-Net) as a valid indicator of lexical complexity (Schmied 2006: 149).

Krhutová (2009) examined the professional discourse of English texts for electrical engineers and lexical cohesive chains in them. The prominent features of lexical cohesion established in this occupational variety were the use of specific terms of electrical engineering (estimated at 30%) and, among members of cohesive chains, it was the use of superordinates.

Lexical cohesion in science and popular science texts was approached from a different prespective than is used in this paper by Myers (1991), who focuses on the relation between the specialised scientific knowledge (domain knowledge) a reader is assumed to have and the explicitly marked cohesive relations. Myers compares scientific and popular articles on one discovery in the field of molecular genetics, claiming that it is possible to create a computational model of cohesion explaining why even a non-specialist reader lacking specialised knowledge can understand scientific texts. He suggests that, "readers of scientific articles must have a knowledge of lexical relations to see the implicit cohesion, while readers of popularizations must see the cohesive relations to infer lexical relations. This difference helps explain some striking differences in the kinds of cohesive devices in the two kinds of texts (…)." (1991: 5) Myers sees the difference in the variety of used cohesive devices, or actually differences in the primacy of lexical vs. grammatical cohesion: "Scientific articles are held together by repetition. Popularizations also depend mainly on repetition, but they can use replacement, conjunctions, pronouns and other devices as well. The fact that the scientific texts do not use pronouns or replacement for cohesion makes them harder for the nonspecialist to follow, whereas the range of devices in the popular texts makes for explicit cohesion that allows the links between sentences to serve as a basis for inferences about the meanings of any unfamiliar terms." (Ibid.)

3.2 Link between cohesion and coherence

As it has been noted above, coherence is achieved by properly applied cohesive devices. Halliday and Hasan (1976: 293) conclude that "the expression of the semantic unity of the text lies in the cohesion among the sentences of which it is composed."

Halliday and Hasan (1989) have later combined the lexical and grammatical cohesion into a single approach defined semantically. Within this approach, cohesive ties, i.e. links between cohesively related items (the term 'cohesive links' denotes cohesive ties holding between two textual items) yield so-called cohesive chains, if such semantic links exist between more than two related items in a text. Cohesive chains may be divided into **identity chains** (where the relation between their members is co-referentiality) and **similarity chains** (the relation between their members is co-classification or co-extension) (Hasan 1989). It is existence of cohesive chains that contributes significantly to achieving coherence of a text (Ibid.).

As for the types of cohesive tie, i.e. links between text items underlying cohesive chains, Hasan (1985) recognizes three types: co-referentiality, co-classification and co-extension. Co-referentiality is based on identity of reference, co-classification is made up of items in an identical class and co-extension is based on a general resemblance. The concept of direct cohesive chains is thus equivalent to the tie of co-referentiality (cf. Zmrzlá 2009: 38, 41-42).

4 Analysis of lexical cohesive devices in texts on physics

4.1 Methods of analysis

The present research is built on the distinction between scientific texts differing in the level of formality, not in the genre as such. However, since genre distinctions are quite significant, an effort was made to include comparable genres of theoretical (or academic) and popular scientific written style in the two parallel corpora. Therefore, the corpus of theoretical scientific texts includes an extract from a university textbook, entries from an Internet encyclopedia and a research article from a printed scientific journal whereas the popular science corpus contains equivalent texts focusing on a more general readership, namely texts from an Internet document containing a popular explanation of physical phenomena, a support material for physics teachers and an article from a popular science webzin. The comparability and a certain representativeness in terms of topics was sought by the choice of texts defining or explaining identical or somehow commensurate areas of physics (see 4.2).

Scientific or academic style and its written mode is quite rich in different genres, e.g. research papers, dissertations and theses, abstracts and résumés/summaries, scientific studies or reports, grant

proposals, etc. Swales (1990: 18-25) also deals with the concepts of genre chains, genre sets and genre networks in the academic/research environment, as different genres are grouped in sets organised by serial order, thematic relatedness, engagement of a particular individual, etc. Scientific texts differ substantially by their specific purpose (referred to as mode in Davy & Crystal 1969), which has been proved by deviations in the use of cohesive devices within each subcorpus in this research.

In agreement e.g. with Zmrzlá (2009) and contrary to Halliday and Hasan´s classic concept from 1976, in my analysis cohesion within a sentence (inter-clausal linkage) is considered relevant, not only cohesion across sentence boundaries. This approach is necessitated by the syntactic complexity of scientific language and texts produced within this variety (cf. Zmrzlá 2009: 39), which leads to the use of cohesive devices marking ties between items in different clauses within complex and compound sentences and thus contributing to the coherence of the whole text.

Also, direct cohesive chains in the present analysis include both reiterations and relexicalisations of the key word of a given text, irrespective of whether they refer directly to the concept in question or to a more general (superordinate) or a more specific (subordinate) concept or whether they are used only as attributes modifying another concept. This concept differs fundamentally from that of Zmrzlá, who only included in her direct cohesive chain the items which referred to the same notion in the extralinguistic context (2009:37). However, not applying the suggested broader approach would result in absence of some basic kinds of lexical cohesion, such as generalisation or specification, as they do not strictly refer to the same entity, but rather to its class or its specific subtype or realisation, respectively. The method used in this paper thus conforms with that outlined in Halliday and Hasan 1976.

Several rules dictated by practical considerations had to be formulated and followed during the analysis:

(1) Lexical cohesion occurs at the same time with grammatical cohesion, so the identification of cohesive chains, analysis and evaluation of cohesion must take into account both types, although the focus then shifts to lexical cohesion, as this is the subject matter of this paper.

(2) Passages of text including numerous physical formulae, equations and calculations have to be avoided or skipped as they do not provide enough linguistic material and cause fragmentation of the syntactic pattern of the text.

(3) Only those expressions are suitable for being chosen as the key expressions of cohesive chains that have a sufficiently concrete or specifiable denotation and that are placed relatively high in a hyponymic hierarchy, i.e. they are neither too general nor too specific, otherwise the established lexical cohesive

relations would be biased qualitatively. Especially suitable are the nominal phrases denoting the key concepts of a given text, occurring in headings of chapters, titles of articles, or serving as encyclopedic entries.

(4) As there is virtually no semantic difference between phrases containing the key notion of a cohesive chain but expressing it via different grammatical structures, e.g. "charge" in "possess charge" and "be charged", both such instances are seen as co-referential and thus members of the same cohesive chain. Word-class distinctions are not so precisely defined in English with its easy conversion and rather a poor repertory of formal markers in the case of using a word in different syntactic roles, so cohesive chains need not be homogeneous in terms of word class.

(5) Lexical and grammatical cohesive devices may be at play at the same time, disabling thus simple classification of a cohesive chain member in question. E.g. *(static) electricity – **the electricity** around you*: it is an example of repetition in terms of lexical cohesion, but the occurrence of the definite article makes it probably also a case of generalisation in contrast with the more specific use of the key expression in the former case(s).

(6) Some types of cohesion overlap and can hardly be separated from each other, or, in other words, their cohesive force combines several distinct types of cohesive means. E.g. two types of lexical cohesion may be difficult to distinguish: *excess positive charge* (spe/rep) – *excess negative charge* (spe/rep) . Is it exact repetition of the focal word, *charge*, modified by some attributes (i.e. a case of collocation), or rather specification (another subtype of reiteration), i.e. a bundle of words which has a different denotation from the single-word key expression and thus a different referent in the real context. The fact that specific (subordinate) items sometimes include the superordinate naming unit as the head word (but sometimes they are lexically different), confuses the matter even more.

(7) Distinction between an open compound lexeme and a collocation is often hard to draw, which leads to an uncertainty whether a pair or a bundle of words are a single lexeme, or just a collocation with the focal word being repeated. E.g. are combinations *like charge* and *unlike charge* two specific subtypes of electric charge, representing a lexical cohesive relation of specification, or are like and *unlike* just freely attached attributes, and the word *charge* is thus directly repeated? And is the pair *positive charge* and *negative charge* more lexicalised (as it was classified in the present paper), and so is to be analysed as the lexical cohesive relation of specification?

(8) Particularly in the theoretical scientific discourse, texts are frequently interrupted by formulae, calculations, visuals (pictures, diagrams, graphs, etc.), examples, subheadings, etc. These are exluded

from the cohesive chains in the present analysis, although they actually extend them, quite often by means of repetition.

4.2 The composition of the corpus and the research questions

The corpus is composed of two parallel subcorpora, the corpus of theoretical (T1-T4) and the corpus of popular (P1-P4) scientific texts. The corpus of theoretical scientific texts (TST) includes the following texts:

T1: "Electric Charge and Matter" (*Physics*, 502-503)

T2: "Electric charge" (*Wikipedia*)

T3: "Friction" (*Wikipedia*)

T4: "On the physical foundations of the method of Sjöstrand for reactivity measurements by the pulsed neutron technique" (*Annals of Nuclear Energy*)

The corpus of popular scientific texts (PST) consists of these texts:

P1: "Electricity" (*Explain that Stuff*)

P2: "Static electricity" (*Explain that Stuff*)

P3: "Friction For Children – 4 Tricks to Help Children Understand Friction" (*Ezine articles*)

P4: "Is the Electromagnetic Radiation from a Laptop Computer Dangerous?" (*Factoidz: bite-sized knowledge*)

Based on the theoretical premises discussed in sections 2 and 3, the research questions were as follows:

Q1: What is the "density" of lexical cohesive chains in the discourse of science (physics here)?

Q2: What is the ratio between devices of lexical and grammatical cohesion in scientific texts (physics)?

Q3: What is the ratio between individual types of lexical cohesion in scientific texts (physics)?

Q4: Are the differences in register (theoretical vs. popular, research vs. didactic) reflected in the use of means of lexical cohesion?

4.3 Corpus of theoretical scientific texts: the analysis

Members of cohesive chains were identified in the individual source texts and tagged by an appropriate abbreviation in brackets. Instances of grammatical cohesion are marked as follows: reference: **(ref)**, substitution: **(sub)** , ellipsis: **(ell)**;

and instances of lexical cohesion have the following abbreviations:
repetition: **(rep)**, equivalence: **(equ)**, generalisation: **(gen)**, specification: **(spe)**, and antonymy: **(ant)**. Cases of multiple class membership are also marked, using a double tag with a slash in between the abbreviations; however, only the first (dominant or less doubted) type counts in the statistics.

Each cohesive chain has been analysed quantitatively to answer questions set in 4.2 and the results have been transferred to a single table to allow comparison. Below is an example of a tagged cohesive chain and the corresponding analysis of types of cohesive devices from the TST corpus:

Friction (rep) – Friction (rep) – the force (gen) *resisting the relative motion of solid surfaces, fluid layers, or material elements sliding against each other – the opposite of* "slipperiness" (ant) – *several types of* friction (rep) – dry friction (spe) – Dry friction (spe) – static friction (spe) *between non-moving surfaces* – kinetic friction (spe) *between moving surfaces* – Fluid friction (spe) – the friction (rep) *between layers within a viscous fluid that are moving relative to each other* – Lubricated friction (spe) – a case of fluid friction (rep/spe) *where a fluid separates two solid surfaces* – Skin friction (spe) – a component of drag (equ) – the force (equ) *resisting the motion of a solid body through the fluid* – Internal friction (spe) – the force (gen) *resisting motion between the elements making up a solid material while it undergoes deformation* – the friction (rep) – friction (rep) – *many types of* friction (rep) – Friction (rep) – a component of the science of tribology (gen) – Friction (rep) – *not a fundamental* force (gen) – friction (rep) – 0 (ell)

Total words: 248 (1 page)
Members of the chain: 27
Chain in % of total words: 27/248 (10.9%)
Grammatical cohesion in the chain: 1/27 (3.7%)
Reference: 0
Substitution: 0
Ellipsis: 1/1 (100%), 1/27 (3.7% of the total)
Lexical cohesion in the chain: 26/27 (96.3%)
Repetition: 11/26 (42.3%), 11/27 (40.7% of the total)
Equivalence: 2/26 (7.7%), 2/27 (7.4% of the total)
Generalization: 4/26 (15.4%), 4/27 (14.8% of the total)
Specification: 8/26 (30.8%), 8/27 (29.6% of the total)
Antonymy: 1/26 (3.8%), 1/27 (3.7% of the total)

Fig. 1: Text and analysis – T3 ("Friction", *Wikipedia*)

4.4 Corpus of popular scientific texts: the analysis

The same procedure of tagging, analysis and calculation of percentages has been used in the PST corpus. As the corpus of popular texts displays a larger heterogeneity, two examples are included to illustrate it. Text P1 seems to be fairly standard within the framework of this (sub)style, with a relatively high proportion of grammatical cohesive items. Text P3 yields different results from "standard" texts in the PST corpus, since it contains no examples of grammatical cohesion in the chain, which is quite nontypical in the PST discourse (unlike the TST discourse, where the proportion of grammatical cohesive devices is distinctively low). Also, text P2 shows deviations, making it difficult to identify the dominant type of lexical cohesion when the key (multi-word) expression is variously paraphrased and/or just one part of the full term is used in the relexicalisations. This results in uncertainty of their classification as either repetitions or equivalents; and when considering equivalents, also whether they are mere relexicalisations of the term (equivalence) or specific types of the entity (specification). Text P2 thus distorts the statistics of the PST corpus, since equivalence has been preferred in the analysis to possible repetition or specification.

Electricity (rep) – electricity (rep) – a bolt of lightning (spe) – *a sudden, massive surge of* electricity (rep) *between the sky and the ground beneath* – The energy in a single lightning bolt (spe) – Electricity (rep) – the most versatile energy source (equ) – it (ref) – one of the newest (sub/ equ) – it (ref) – Electricity (rep) – it (ref) – own renewable electric power (equ/spe) – electricity (rep) – it (ref) - **electricity** (rep) – Electricity (rep) – a type of energy (gen) – that (ref) – 0 (ell) – electricity (rep) – it (ref) – static electricity (spe) – electricity (rep) – that (ref) – current electricity (spe)

Total words: 210 (1/2 a page)

Members of the chain: 26

Chain in % of total words: 26/210 (12.4 %)

Grammatical cohesion in the chain: 9/26 (34.6%)

Reference: 7/9 (77.8%), 7/26 (26.9% of the total)

Substitution: 1/9 (11.1%), 1/26 (3.8% of the total)

Ellipsis: 1/9 (11.1%), 1/26 (3.8% of the total)

Lexical cohesion in the chain: 17/26 (65.4%)

Repetition: 10/17 (58.8%), 10/26 (38.5% of the total)

Equivalence: 2/17 (11.8%), 2/26 (7.7% of the total)

Generalization: 1/17 (5.9%), 1/26 (3.8% of the total)

Specification: 4/17 (23.5%), 4/26 (15.4 % of the total)

Antonymy: 0

Fig. 2: Text and analysis – P1 (“Electricity“, *Explain that Stuff*)

Friction (rep) – **Friction** (rep) – friction (rep) – friction (rep) – this concept (gen) – friction (rep) – fiction (rep) – Friction (rep) – Friction (rep) – Friction (rep) – a push (equ) – a pull (equ) – a force (gen) *which works against the motion of objects that are in contact as they move past each other* – friction (rep) – friction (rep) – *three types of* friction (rep) – sliding friction (spe) – rolling friction (spe) – fluid friction (spe) – Sliding friction (spe) – sliding friction (spe) – Rolling friction (spe) – rolling friction (spe) – Rolling friction (spe) – sliding friction (spe) – fluid friction (spe) – fluid friction (spe) – the friction (rep) – drag (equ) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – fiction (rep) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – friction (rep) – the friction (rep) – *reduce* friction (rep)

Total words: 641 (1 page)

Members of the chain: 44

Chain in % of total words: 44/641 (6.9%)

Grammatical cohesion in the chain: 0/44 (0%)

Reference: 0

Substitution: 0

Ellipsis: 0

Lexical cohesion in the chain: 44/44 (100%)

Repetition: 28/44 (63.6%), 28/44 (63.6% of the total)

Equivalence: 3/44 (6.8%), 3/44 (6.8% of the total)

Generalization: 2/44 (4.5%), 2/44 (4.5% of the total)

Specification: 11/44 (25%), 11/44 (25% of the total)

Antonymy: 0

Fig. 3: Text and analysis – P3 (Moffat, L., “Friction For Children – 4 Tricks to Help Children Understand Friction“, *Ezine articles*)

4.5 Comparison of frequency of lexical cohesive devices

It is interesting to compare the results of the analysis of lexical cohesive devices in the theoretical vs. popular physics discourse with findings of a similarly focused study, which was

conducted by Zmrzlá (2009) on a single scientific article from the field of computer science. Zmrzlá observed the features of both grammatical and lexical cohesion as they were manifested in the direct cohesive chain consisting of items referring to the key expression and the main topic described in the article. Particularly the ratio between grammatical and lexical cohesive devices is worth mentioning and comparing with the findings of the present study: Zmrzlá identified more than three times as many aspects of lexical cohesion than those of grammatical cohesion in the cohesive chain (2009: 47-48). Also, by far the most frequent type of lexical cohesion was repetition (53.5% of all members of the cohesive chain), followed by generalisation (18.6%). Grammatical cohesion was quite evenly divided between ellipsis (12.8% of the items of the cohesive chain) and reference (10.5%). No instance of substitution and specification were found (Ibid.), which was explained by specifics of the topic in question. In the present research, particularly the results for repetition in the TST corpus oscillate around the figure given by Zmrzlá, and the results for reference and ellipsis prove their regular occurrence in the TST discourse, though the figures are lower in the physics corpus.

Source	T1 *Physics*	T2 *Wiki-pedia*	T3 *Wiki-pedia*	T4 *Annals of Nuclear Energy*	P1 *Explain that Stuff*	P2 *Explain that Stuff*	P3 *Ezine articles*	P4 *Factoidz*
Focal word(s)	electric charge	electric charge	friction	the method of Sjöstrand for reactivity measure-ment	electricity	static electricity + electric charge	friction	electro-magnetic radiation + EMF radiation
Total words	1,065	778	248	415	210	509	641	572
Words in the cohesive chain	**49**	**77**	**27**	**18**	**26**	**40**	**44**	**30**
Chain in % of total words	4.6%	9.9%	10.9%	4.3%	12.4%	7.9%	6.9%	5.2%
Grammatical cohesion (GC)	**5/49**	**8/77**	**1/27**	**3/18**	**9/26**	**5/40**	**0/44**	**8/30**
GC as % of the chain	**10.2%**	**10.4%**	**3.7%**	**16.7%**	**34.6%**	**12.5%**	**0%**	**26.7%**
Reference (% of GC / % of the chain)	20% / 2%	62.5% / 6.5%	0	100% / 16.7%	77.8% / 26.9 %	80% / 10%	0	100% / 26.7%
Substitution (% of GC / % of the chain)	40% / 4.1%	25% / 2.6 %	0	0	11.1% / 3.8%	0	0	0
Ellipsis (% of GC / % of the chain)	40% / 4.1%	12.5% / 1.3%	100% / 3.7%	0	11.1% / 3.8%	20% / 2.5%	0	0
Lexical cohesion	**44/49**	**69/77**	**26/27**	**15/18**	**17/26**	**35/40**	**44/44**	**22/30**
% of the chain	**89.8%**	**89.6%**	**96.3%**	**83.3%**	**65.4%**	**87.5%**	**100%**	**73.3%**
Repetition (% of LC / % of the chain)	79.5% / 71.4%	65.2% / 58.4%	42.3% / 0.7%	13.3% / 11.1%	58.8% / 38.5%	17.1% / 15%	63.6% / 63.6%	27.3% / 20%

Equivalence (% of LC / % of the chain)	2.3% / 2%	2.9% / 2.6%	7.7% / 7.4%	53.3% / 44.4%	11.8% / 7.7%	60% / 52.5%	6.8% / 6.8%	54.5% / 40%
Generalisation (% of LC / % of the ch.)	2.3% / 2%	5.8% / 5.2%	15.4% / 4.8%	13.3% / 11.1%	5.9% / 3.8%	11.4% / 10%	4.5% / 4.5%	4.5% / 3.3%
Specification (% of LC / % of the chain)	9.1% / 8.2%	26.1% / 23.4%	30.8% / 9.6%	20% / 16.7%	23.5% / 15.4%	11.4% / 10%	25% / 25%	13.6% / 10%
Antonymy (% of LC / % of the chain)	6.8% / 6.1%	0	3.8% / 3.7%	0	0	0	0	0

Table 1: Summary of the proportions of cohesive ties used in the corpora of theoretical scientific texts (T1-T4) and popular scientific texts (P1-P4).

5 Conclusions

The analysis of individual texts in both corpora has showed that lexical cohesive chains with the key word make from 4.3% to12.4% of the total wordcount (in fact even more, as the multi-word character of chain members was not taken into consideration). No obvious TST/PST difference has been observed.

Lexical cohesive devices in the chains accounted for 65.4%-100%, on average nearly 90% in the TST corpus, over 80% in the PST corpus (and even less without the distortion caused by text P3). Conversely, grammatical cohesion is more frequent in the PST chains (with one 0% exception), its most prominent type being reference.

The most frequent type of lexical cohesive tie is repetition (including derivatives), more frequent in the TST corpus (between ca. 40-70% of the whole chain, except text T4 taken from a journal). Equivalence seems to be more frequent in the texts of the PST corpus, sometimes replacing repetition as the dominant lexical cohesive tie (texts T4, P2, P4). Its occurrence is clearly reciprocal to the more exact repetition.

Specification is the second most frequent type of lexical cohesive tie, not revealing a substantial TST/PST difference. Lastly, generalisation is marginally a more frequent type in the TST corpus, which may be connected with more references to hierarchical relations (the same applies to a higher use of specification, though just in some of the theoretical scientific texts).

References

Crystal, D. and Davy, D. (1969) *Investigating English Style*. London: Longman.

Dontcheva-Navratilova, O. (2006) 'Cohesion in colony texts.' In Drápela, M. and Vomlela, J. (Eds.) *Silesian Studies in English 2006: International Conference of English and American Studies, Opava 18-19 September 2006* (Proceedings). Opava: Slezská univerzita v Opavě. 51-61.

Dontcheva-Navratilova, O. (2005) *Grammatical Structures in English: Meaning in Context.* Brno: Masarykova univerzita.
Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English.* London and New York: Longman.
Halliday, M.A.K. and Hasan, R. (1989) *Language, Context and Text: Aspects of Language in a Social Semiotic Perspective*. 2nd ed. Oxford: OUP.
Hasan, R. (1985) 'The texture of a text.' In Halliday, M.A.K. and Hasan, R. *Language, Context and Text: Aspects of Language in a Social Semiotic Perspective*. 2nd ed. Oxford: OUP, 70-96.
Knittlová, D. (1990) *Funkční styly v angličtině a češtině.* Olomouc: Univerzita Palackého, filozofická fakulta.
Krhutová, M. (2009) 'Cohesion and coherence in written texts for professional and academic purposes.' In Dontcheva-Navratilova, O. and Povolná, R. *Coherence and Cohesion in Spoken and Written Discourse*. Newcastle upon Tyne: Cambridge Scholars Publishing. 154-166.
Myers, G. (1991) 'Lexical cohesion and specialized knowledge in science and popular science texts.' *Discourse Processes 14:1*, 1-26.
Schmied, J. (2006) 'Specialist vs. non-specialist academic discourse: Measuring complexity in lexicon and syntax.' In Povolná, R. and Dontcheva-Navratilova, O. *Discourse and Interaction 2*. Brno: Masarykova univerzita. 143-152.
Swales, J.M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
Urbanová, L. (2008) *Stylistika anglického jazyka*. Brno: Společnost pro odbornou literaturu – Barrister & Principal.
Wales, K. (2001) *A Dictionary of Stylistics.* 2nd edition. Harlow: Pearson Education Limited.
Zmrzlá, P. (2009) 'Direct cohesive chains in scientific articles.' In Hůlková, I. and Jančaříková, R. (Eds.) *Exploring Cohesion and Coherence in English Discourse. Third Brno Conference on Linguistic Studies in English.* Brno: Masarykova univerzita. 37-50.

Sources

'Electric charge'. *Wikipedia, the free encyclopedia*. Retrieved 9 Sept. 2010 from http://en.wikipedia.org/wiki/Electric_charge.
'Electricity.' *Explain that Stuff.* Retrieved 9 Sept. 2010 from http://www.explainthatstuff.com/electricity.html.
'Friction'. *Wikipedia, the free encyclopedia*. Retrieved 9 Sept. 2010 from http://en.wikipedia.org/wiki/Friction.
Isbasescu, M. (1977) 'On the physical foundations of the method of Sjöstrand for reactivity measurement by the pulsed neutron technique'. *Annals of Nuclear Energy 4*, 193-200.
Keller, F.J., Gettys, W.E., and Skove, M.J. (1993) *Physics*. Second edition (1st ed. 1989).
Kramer, M. 'Is the Electromagnetic Radiation from a Laptop Computer Dangerous?' *Factoidz: bite-sized knowledge*. Retrieved 13 Sept. 2010 from http://factoidz.com/is-the-electromagnetic-radiation-from-a-laptop-computer-dangerous/.
Moffat, L. 'Friction For Children – 4 Tricks to Help Children Understand Friction.' *Ezine articles.* Retrieved 14 Sept. 2010 from http://ezinearticles.com/?Friction-For-Children.
'Static Electricity.' *Explain that Stuff.* Retrieved 9 Sept. 2010 from http://www.explainthatstuff.com/electricity.html.