LEXICAL PATTERNING IN RESEARCH ARTICLES: TOWARDS A WORKABLE SUMMARY

Gabriela Zapletalová

Abstract

The contribution aims to test the methodology based on the concept of lexical cohesion referred to as *lexical patterning* by Hoey (1991) who proposed the model with the aim to decide (1) how lexis organises text and (2) introduced a framework based on the analysis of lexical patterning for the production of intelligible abridgements or summaries of non-narrative texts. Lexical patterns will be investigated in the genre of argumentative research article that has been chosen as a model type of non-narrative text. The method for revealing lexical patterning has been proposed with the aim to extract essential information and present it in a condensed form that would serve as a kind of comprehensive abstract of the whole text. The reason for producing readable summaries using Hoey's methodology is quite reasonable and relevant: it is an enormous growth of various types of information sources: there are hundreds of research articles (RAs) that appear every year and it is beyond one's own limits to access all texts and absorb all information presented. We will test to what extent the resulting summary is workable in the way that it provides the reader with the information whether the original text is worth reading.

Key words

lexical cohesion, lexical patterning, research article, summary, text

1 Introduction

The reason for producing workable summaries is quite well-founded and relevant: it is an enormous growth of various types of information sources. In the field of academic writing there are hundreds of RAs that appear every year and it is beyond one's own limits to access all texts and absorb all information presented. The method of lexical patterning has been proposed with the aim to extract essential information and present it in a condensed form that would serve as a kind of comprehensive 'abstract' of the whole text. The resulting summary should be reliable in the way that it provides the reader with the information whether the original text is worth reading: "a [good] summary acts as a filter, indicating the major content of the original story" (de Oliveira et al. 2002: 1). The paper aims at testing the methodology based on the concept of lexical cohesion (LC) referred to as *lexical patterning* by Hoey (1991) who proposed it with the aim to decide (1) how lexis organises text and consequently, (2) introduced a framework based on the analysis of lexical patterning for the production of intelligible abridgements or summaries of non-narrative texts.

2 Lexical cohesion

One of still most influential studies on cohesion is Halliday and Hasan's (1976) *Cohesion in English*: cohesion is a semantic concept realised through the lexico-grammatical system; it reflects the relations of meaning that exist within the text and define it as a text; it is a surface phenomenon based on non-structural, text-forming relations. Hoey's (1991) account of cohesion underlies its character of an objective property of the text. He stresses its organisational and 'interpretative' quality: it is "the way certain words or grammatical features of a sentence connect that sentence to its predecessors (and successors) in a text" (1991: 3). The view of a text as organised through cohesion rather than a structured unit is close to Halliday's (1994) idea of cohesion as a processual relation between entities.

Halliday and Hasan treat LC as "selecting the same lexical item twice, or selecting two that are closely related" (1976: 12). They introduce two basic concepts: (i) a *tie* as a "single instance of cohesion", and (ii) *texture* which is explained as the property of "being a text" (1976: 2-3). The organisation of text (*texture*) is formed by relationships that exist among items of text. These relationships are either grammatical or semantic and they create *cohesive ties*. Hoey's (1991) methodology inspired by Halliday and Hasan (1976), Hasan (1984) and Stotsky (1983) is based on the assumption that "lexical cohesion is the only type of cohesion that regularly forms multiple relationships" (1991: 10). Hoey's LC rests on a simple presupposition suggested by Halliday and Hasan (1976: 292): "however luxuriant the grammatical cohesion displayed by any piece of discourse, it will not form a text unless this is matched by cohesive patterning of a lexical kind". In brief, LC is a necessary prerequisite for creating texture: "the study of cohesion (...) is to a considerable degree the study of patterns of lexis in text" (Hoey 1991: 10). Thus lexis has been assigned leading role in the construction of text and text's organisation.

Central is the category of *repetition*. The term repetition may induce the mistaken idea that repeating the same lexical item counts among monotonous and stereotypical ways of expression. I must agree with Tárnyiková (2002) who admits (with respect to narrative texts she focuses on) that it is not a very creative way of text-shaping, but at the same time stresses its special communicative value even in narrative texts. Non-narrative (scientific) texts avoid creative intricacies so as to preserve exactness and terminological preciseness.

Hoey (1991) classifies repetition into four main lexical types (1–4) and four minor classes (5–8), as in Table 1. *Simple/complex paraphrase* and *superordinate/hyponymic repetition* are lexically semantic classes expressing various degrees *of*

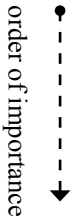
semantic contiguity (Tárnyiková 2002): synonymy, antonymy, or marking other lexical and sense-relations.

1. Simple repetition	two identical items (*volume – volume*) or two similar items whose difference is 'entirely explicable in terms of a closed grammatical paradigm' (*volume – volumes; indicate – indicated*) (Hoey 1991: 53).
2. Complex repetition	two lexical items sharing a lexical morpheme but differing with respect to other morphemes (*drug – drugging, productivity – production; audible – inaudible*) or grammatical functions (*humans* (N) – *human* (Adj); *write – writer*.
3. Simple paraphrase	two different lexical items of the same grammatical class when one item "may substitute for another in context without loss or gain in specifity and with no discernible change in meaning" (Hoey 1991: 62) (e.g. *traditional – standard, expansion – growth, explain – interpreted*: also changes in the grammatical paradigm). It is an attempt at "faithful but autonomous restatement" (Wales 2001: 284).
4. Complex paraphrase	two different items of the same or different grammatical class; it occurs when "two lexical items are definable such that one of the items includes the other, although they share no lexical morpheme" (Hoey 1991: 64): a) antonyms without a lexical morpheme (*growth – decline*) b) one item is a complex repetition of another (*decline – declines*) and also a simple paraphrase (or an antonym) of a third (*decline – growth*) c) it is possible to substitute an item for another: a complex paraphrase between *record* and *discotheque* if *record* can be replaced with *disc* (Hoey 1991: 66).
5. Hypernymic repetition	one lexical item is followed by the more general – superordinate – word such as in e.g. *horsepower → power unit*, *technician → the expert*.
6. Hyponymic repetition	the latter word does not include the former; it is based on reverse order principle: *general word → specific word*.
7. Co-reference repetition	it has been employed in order to clarify the doubtful cases of superordinate and hyponymic repetition where the shared context cannot be taken as a reliable guide such as in the case of *Tony Blair → the Prime Minister* or *scientists → biologists*, in which the only decisive feature can be the common referent.
8. Substitution	the category that subsumes other ways of repeating such as reference, substitution and ellipsis in Halliday and Hasan's (1976) account.

Table 1: Lexical cohesion by Hoey (1991)

Out of the eight lexical repetition categories mentioned above, (1–4) are treated as most significant to our purposes because they represent lexical items the meaning of which is definitional, and thus they do not depend on other items for their interpretation. In addition, the aim of the analysis is to study lexical links that contribute to the patterning effect in the text. The minor categories are based on lexical links as well, but their identification and interpretation is more complex due to the criterion of the common referent as the critical feature.

Substitution is regarded as a minor category because it is based on the repetition of textual items – grammatical links – that also contribute to repetition, but is usually employed for the purpose of local cohesion. The textual items that include the pronoun system and substitutes depend on other items for their interpretation. All eight categories can be arranged from the most to least significant for the purposes of analysis. The following hierarchy of repetition classes is based on Hoey's (1991) classification:



Simple repetition
Complex repetition
Simple mutual paraphrase
Simple partial paraphrase
Complex paraphrase (antonymous and other paraphrase)
Co-reference repetition (superordinate repetition and hyponymic repetition)
Substitution

The decreasing order of importance does not imply that for instance *substitution* in general plays a less significant role in text construction than *simple repetition*. Rather it has much to do with our subjective ranking of importance for the purposes of the present analysis, in which lexical relations are given preference to grammatical ones.

The most difficult and crucial task in this analysis is to identify the particular kind of repetition. For this reason, two key concepts are introduced, *links* and *bonds*, together with *repetition matrices*, which are proposed in order to help establish the number of connections between sentences. A *link* is identified on the basis of endophoric reference (in particular anaphoric reference in the data) according to which the link shows a relation between two lexical items in two separate sentences or sentence complexes. Therefore, repetition links that occur in one sentence are not recorded. The concept of *bond* has been proposed with the aim of measuring relations between sentences. Bonds are established when two sentences are linked by a certain number of links, or as Hoey (1991: 91) puts it, a bond arises as "a connection made between any two sentences by virtue of there being a sufficient number of links between them". In his account, the number of links suggested as creating a bond is three (to reduce so-called chance cohesion). In texts where most sentences are connected by three and more links, the number for creating a bond will be necessarily higher. Still, three links remains the lowest number for establishing a bond "because of the greater

likelihood of two repetitions occurring in a pair of sentences by chance or as a way of characterizing a single topic" (ibid.: 190). The concept of links and bonds is shown in Figure 1, which records two types of links that are responsible for what may be called local and distant (or long-distance) cohesion.

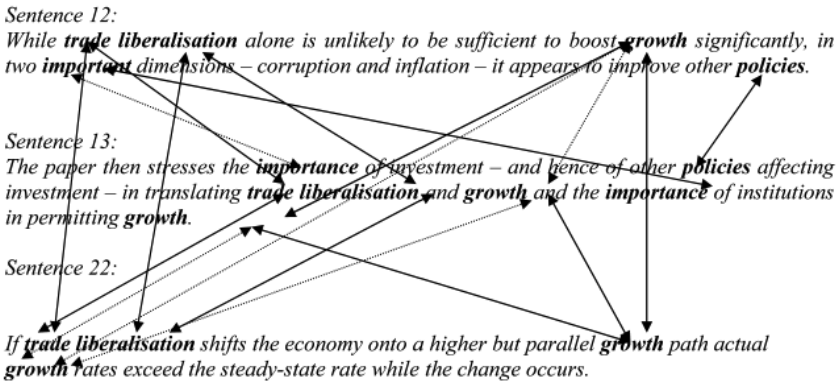


Figure 1: Bonded sentences (from Winters, EJ, p. 4)

Inevitably, the quantity of bonding is text-dependent; it must be adjusted according to its type. In Figure 1, the relatively strong bond between sentence (12) and sentence (13) is established by five links, and there is a three-link bond between (12) and (22), and between (13) and (22). This example also illustrates two double links (a broken line) between one occurrence of *growth* in (12) and two in (13) and (22); similarly *important – importance*, but in this case only one link is established. Again, the relation between (12) and (13) with (22) can be treated as an example of distance bonding operating over long stretches of text.

Sentence 0	(-, 1)
Sentence 1	(0, 0)
Sentence 3	(0, 0)
Sentence 4	(0, 2)
Sentence 5	(0, 0)
Sentence 6	(2, 0)
Sentence 7	(1, -)

Table 2: Record of bonding derived from the repetition matrix

Having identified links and bonds, we arrive at a quantification of the cohesive quality of all sentences in the text, which can be demonstrated as a coordinate: the first number shows references to previous sentences; the second shows references to following ones: Table 2 shows the closeness of the connection between sentences: a two-figure coordinate reveals the number of references (by way of example, a bond is formed by a minimum of three links) of each sentence to another.

3 Research methodology

Hoey tested his concept on a 40-sentence introductory section from a textbook on philosophy. Since our study focuses on RAs we have decided to test lexical patterning on the whole texts to find out whether the concept can be utilised as a method for creating intelligible summaries out of large-scale non-narrative texts. The final choice was for two complete texts including 10,451 running words.

Hypothesis 1: *Marginal* sentences manifest no bonds in the text: having lower information value, they do not contribute to a topic of the text; the information they carry is referred to as metatextual. *Marginal* does not mean unimportant.

Hypothesis 2: *Central* sentences manifest more bonds in the text. We expect that RAs will reveal higher cut-off point than three; however, the number must be set individually. The extraction of central sentences will enable us to create a meaningful summary.

Hypothesis 3: *Topic-controlling* sentences may be important for the creation of a summary because they are responsible for the development and control of the topic. They are *topic-opening*, i.e. those that have bonds with successive sentences – their second coordinate is higher (e.g. 2, 15); and *topic-closing* which have bonds with previous sentences (e.g. 12, 3).

Hypothesis 4: Regarding the properties and nature of the bonds, we can test what Hoey (1991: 125) calls the *strength of bonds*. It is based on the closeness of bonded sentences: it was found out that bonds formed between adjacent sentences represent a relatively small number. Therefore the closeness of non-adjacent bonded sentences will be measured through the weak and strong claim: *The weak claim*: each bond marks a pair of sentences that is semantically related in a manner not entirely accounted for in terms of its shared lexis. *The strong claim*: because of the semantic relation referred to in the weak claim, each bond forms an intelligible pair in its context.

Hoey's concept has been slightly modified for the purposes of the present analysis in order to capture the thematic centrality of particular sentences not only within each chapter of the text but also across the chapter-boundary: RAs consist of thematically and formally precisely planned chapters that unfold as

independent distinct units. First, individual chapters were analysed as separate texts; second, sentences from the introductory section with the highest number of bonds, which may be measured as the signal of their thematic centrality, have been used in the analysis of the following chapter and so on in the remaining text. Then potential summaries were constructed from the categories of marginal, central and topic-controlling sentences. Within each category the sentences were ranked according to the number of bonds each sentence has with other remaining sentences in the text.

One interesting aspect of our analysis has been to set the cut-off point for the constitutive number of bonds since various chapters of the RAs revealed different results. The minimum number of bonds was set for five; the maximum possible number was eight in one chapter: therefore seven bonds were decided upon as the maximum number. Consequently, three resulting summaries were constructed as shown in Table 3.

Type of summary based on:	**Text 1**	**Text 2**	**Average**
Marginal sentences	88 = 32.9%	50.5 = 30%	69 = 31.4%
Central sentences	80 = 30%	44 = 26.2%	62 = 28.1%
Topic-controlling sentences	47 = 17.5%	43 = 25.6%	45 = 21.6%

Table 3: The summaries chosen for the analysis

4 Summary evaluation and interpretation

The process of summary evaluation is a complex one since it draws on two types of data: quantitative and qualitative. Quantitative data are shown in Table 3 which also presents the avegared results of the original texts; the percentage fluctuates from 31.4 as the highest figure to 21.6 as the lowest. However, the quantitative analysis must be accompanied by a proper qualitative evaluation. Qualitative evaluation is based on a subjective judgement; to avoid or rather minimise subjective process we agreed upon three evaluation criteria (cf. de Oliveira et al. 2002: 5):

1. the inclusion of the essential information
2. the exclusion of non-essential information
3. the readability of the summary

Since RAs rank among scientific text types whose aim is to present precise, matter-of-fact and relevant information in a way that does not allow ambiguous interpretation, seminal for our analysis is the inclusion of the essential information.

Marginal sentences: 31.4 per cent

This summary represents the most consistent and coherent text out of the three summaries evaluated: it subsumes the relevant information needed for understanding the gist of the topic. Slightly problematic are (i) deictic pronominals (*this*, *these: a good example of* <u>*this*</u> in (1) and (2); and (ii) the expressions that subsume various types of adverbials: adjuncts and conjuncts*: then*, *previously*, *despite, instead* (Quirk et al. 1985).

(1) ***5*** *The initial impact of a GPT [General Purpose Technologies] on overall productivity growth is typically minimal and the realisation of its eventual potential may take several decades such that the largest growth effects are quite long-delayed, as with electricity in the early twentieth century (David, 1991).* ***8*** *A good example of* ~~*this*~~ ***[a growth slowdown]*** *is taken by the GPT literature to be the hiatus between steam and electricity in the later nineteenth century (Lipsey et al., 1998b), echoing the famous hypothesis first advanced by Phelps-Brown ....*

(2) ***21*** *Section 3 builds on* ~~*these data*~~ ***[= the diffusion of steam power in Britain between 1760 and 1910]*** *to provide growth accounting estimates of the contribution of steam to labour productivity growth,* ***[<u>uses these to address the questions posed in this introduction</u>]*** *and offers some reflections on the elusive issue of TFP spillovers.*

(3) ***74*** *From this it is possible to derive optimal labour supply L* = L* (α, B); which is assumed to be non-decreasing in α as before.* ***79 Hence*** *we cannot obtain clear predictions about how self-employed labour supply responds to a MPS in θ.*

Most frequent are listing, resultative and summative conjuncts *then, hence* and *thus* (3). Both deictics and adverbials were removed if possible as unwanted cohesion (crossed) and replaced by appropriate information (in bold) as in sent. 60 (Text 1) where the improvement of the original resulted in adding the missing equation *(3)* **as in (4)**:

(4) ***60*** *The right-hand side of* ~~***(3)***~~ ***[αЄ[u'(.)] + v'(.) = 0, is positive if u'"(.) > 0]****, in which case A(θ) is convex, and labour supply increases with the level of uncertainty.*

The summaries contain a number of in-text references such as *derived below*, *see below*, or *see Table 1* that are employed in order to 'instruct' the reader. Some phrases had to be omitted because they were co-referential with an item that was not included in the summary such as in sent. 25 (Text 1):

(5) ***25*** *But* recent *research findings (Blundell and MaCurdy, 1999,* ***Table 1****),* ~~*as well as some new research findings below*~~*, cast doubt on the relevance of a backward bend in employee labour supply.*

Another problem for generating the summary arises from numerous equations, tables and graphs that accompany economics texts. In Text 2 all the eight tables were included in the summary since it contained the sentences with references to the tables. More complex is the inclusion of equations into the text as in sent. 34 (Text 1) which is referred to later in the text but does not show any logical link with sent. 46. Still it is a highly bonded sentence as it appears in all summaries based on Text 1. The 'solitariness' of sent. 34 lies in the fact that it is also an important topic-opening sentence opening the first paragraph of the theoretical chapter and together with the equation it contains all seminal elements developed later:

(6) ***34*** *Consider a set of identical self-employed individuals who work alone, and earn income* $y = g(wL^* + B; \theta)$, *(1)*
where w is the hourly wage, L^* *is the chosen number of hours worked* ***(derived below)****, B is unearned income and* θ *is a random variable capturing uncertainty.* ***46*** *To derive our results on self-employed labour supply, we propose a simple model of self-employment income generation that leads to two different specifications of (1).*

Central sentences: 28.1 per cent

This summary including 28.1 per cent of the original version is very close to the previous summary. Similarly, most problematic are the conjuncts *hence* and *then* that do not enable us to compensate for the missing information easily since these sentence connectors contain an enormous amount of contextual information.

(7) ***60*** *The right-hand side of* ~~***(3)***~~ ***[A"(θ) = α u"'(.)]*** *is positive if u'"(.) > 0, in which case A(θ) is convex, and labour supply increases with the level of uncertainty.* ***62*** *Hence* ~~*this*~~ ***[labour supply with the level of uncertainty]*** *'usual' case has the self-employed 'self-insuring' in response to greater uncertainty, choosing a larger labour supply and thereby making the deterministic part of their income relatively larger.*

Topic-controlling sentences: 21.6 per cent

The summary is based on topic-opening and closing sentences that appear at the beginning or end of the paragraph, respectively. Nevertheless, only approx. 41 per cent of all sentences in the summary are real topic-opening and closing. The remaining sentences most usually either immediatelly follow or precede the topic-opening or closing sentences. Text 2 is more consistent and coherent in contrast to Text 1 which is most deficient in chapters 1 and 2 to the extent that we miss the main outline.

Testing strong and weak claims

The last step is to assess the quality and density of bonding through strong and weak claims. The most bonded is sent. 34 in Text 1; let us test its bonding quality with the most distant sent. 177 (= 143 sentences between both):

(8) ***34*** *Consider a set of identical self-employed individuals who work alone, and earn income* $y = g(wL^* + B; \theta)$, *(1)*
where w is the hourly wage, L is the chosen number of hours worked (derived below), B is unearned income and θ is a random variable capturing uncertainty.* ***177*** *As a preliminary check for possible selection bias, we re-computed Table 1 for all self-employed individuals with positive incomes and work hours.*

The relation between 34 and 177 seems to be rather weak than strong since we would need more accompanying text to make the pair contextually cohesive. Similar is the situation with the pairs 34 and 116 (the distance 82 sentences), or 121 and 140 (19 sentences). However, in the case of 34 + 116 we are prone to qualify the bonding as rather strong than weak since they are related through wider context of economic text in which it is not unusual to meet texts consisting of freely consecutive stretches of text having enumerative-like character.

(9) ***34****. viz (8)* ***116*** *Let* L_{it} *be the observed number of hours supplied by a self-employed individual i at t; let* w_{it} *be their computed hourly wage (which, as this is a self-employed sample, we will discuss shortly); and let* r_{it} *be a measure of wage uncertainty.*

The pair 121 + 140 is linked very heavily but the resulting effect is redundant: almost the same or partly overlapping information is introduced which results in clumsy effect on the reader. This is due to the fact that both sentences appear in different sections of the RA.

(10) ***121*** *While different measures of uncertainty could be proposed, we choose to measure* r_{it} *as the standard deviation of i's log wages in contiguous time periods prior to t.* ***140*** *Because* r_{it} *is measured as the standard deviation of wages observed over* n_{it} *periods – where* n_{it} *can vary from individual to individual – at least two previous observations on each i's earnings are needed to compute* r_{it}.

The last two pairs are the examples of good strong claim which means that they make sense without any supplementary text. This speaks for a high lexical density and quality of distant bonding which is in these instances 131 and 135 sentences.

(12) ***34*** *viz (8)* ***165*** *The sample was restricted to self-employed individuals who earned positive incomes and worked positive numbers of hours (but no greater than 4,000 annually), who had at least three consecutive years of income data, and single continuous spells in self-employment (the latter to avoid contaminating estimates of the effects of risk with occupational switching).*

(13) ***34*** (above) ***169*** *Hourly wages were computed by dividing annual earned income by annual hours worked.*

5 Conclusion

The aim was to test Hoey's framework for generating readable summaries out of non-narrative texts. Out of the three procedures based on (1) omitting marginal sentences, (2) including central, and (3) topic-controlling sentences the results are as follows:

Hypothesis 1: The summary based on omitting marginal sentences is relatively very workable despite the fact that the summary is based on a bond defined by five links 'only'. It has been confirmed that marginal sentences have low information value.

Hypothesis 2: Central sentences form a highly coherent summary; the cut-off point for setting them was one bond defined by six links: the summary contains 28.1 per cent of the original which seems to be a suitable amount. However, the summary is similar to the previous one.

Hypothesis 3: Topic-controlling sentences in the data do not appear where they are expected to be: there is a strong tendency of these sentences to appear rather in the middle or immediately after or before topic-opening/closing sentences. Our results are similar to Vašáková's (2006) who analysed science popularisations in *National Geographic*.

Hypothesis 4: Testing strong and weak claims confirms rather good quality of lexical density which is seminal for distant bonding: in most cases the distance around 80 sentences between two lexically-bound units is not problematic.

Overall, all the summaries contain a constant number of sentences that are responsible for the *core of the text*. We can claim that Hoey's framework is essentially practical for the purposes of generating intelligible summaries. The analysis has confirmed that the functionally optimal cut-off point for RAs is as high as six, even seven links in some cases (cf. Hoey 1991: 92). Regarding the number of sentences necessary for such a summary we have arrived at 31.4 per cent which is in harmony with de Oliveira et al. (2002) who claim for around 3 per cent of the size of the original text.

Despite the fact that the results of the analysis fully support Hoey's concept about the organisation of the text through lexical cohesion we are aware of the fact that Hoey manifested his theory on philosophical texts that represent a genre disparate mainly from economic texts in the data. The specifics of economic texts lie in their interconnection with instrumental devices such as equations, graphs, charts, tables and statistics, i.e. textual units with metadiscursive labelling function. Most equations are accompanied by 'explanatory' sections where various subscripts and indexes are explained. As a result, the parts containing most equations – often theoretically-oriented chapters – are the least cohesive. In contrast, most cohesive parts are introductory sections since these concisely and lucidly outline the RA's content. Most problematic seem to be theoretically-biased chapters 1 and 2 that present theoretical background and feedback for own research. However, this disproportion is fully compensated for in the texts through introductory section and then by those conclusive chapters 3–5. This confirms and reflects carefully planned and organised layout of RAs.

One more important point arises when considering how demanding and laborious task it is to construct such a summary since there are some computer tools such as an electronic-like resource WordNet, SummariserPort and other summary-generating systems based on lexical cohesion. However, these tools are able to represent a computer implementation only of the first two of Hoey's four categories of lexical patterning, i.e. *simple* and *complex repetition*. Hence, *simple* and *complex paraphrase* classes are not included since paraphrases are not based on closed class words and various non-lexical items that could be easily encoded into the programme. In this case, the identification of these relations in texts underlies the existence of so-called sophisticated lexicon that would help identify e.g. the co-referential link between two lexically unrelated items *the Prime Minister* → *Tony Blair* where the context is important: the link would be established only for the period of 1997-2007 when Blair actually held the

post. In contrast, the present analysis made use of all repetitive classes including paraphrases which can be done so far only manually. Moreover, the research was meant as a probe into the nature and characteristics of links and bonds in economics texts and their role in capturing thematic centrality of specific sentences in the texts.

References

Halliday, M. A. K. (1994) *Introduction to Functional Linguistics.* London: Arnold.
Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London: Longman.
Hasan, R. (1984) 'Coherence and cohesive harmony.' In: Flood J. (ed.) *Understanding Reading Comprehension*. Delaware: International Reading Association. 181-219.
Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.
de Oliveira, P. C. F., Ahmad, K. and **Gillam, L.** (2002) '**A financial news summarization system based on lexical cohesion.'** In: Gillam L. (ed.) *Terminology and Knowledge Engineering: Making Money in the Financial Services Industry.* Proceedings of a Workshop at the International Conference on Terminology and Knowledge Engineering, 30 August 2002, Nancy France. 1-6.
Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language.* London: Longman.
Stotsky, S. (1983) 'Types of lexical cohesion in expository writing: Implications for developing the vocabulary of academic discourse.' *College Composition and Writing, 34*(4), 430-445.
Tárnyiková, J. (2002) *From Text to Texture. An Introduction to Processing Strategies.* Olomouc: Univerzita Palackého.
Vašáková, V. (2006) *Lexikální koheze v textech časopisu* National Geographic Magazine [Lexical cohesion in texts from "National Geographic Magazine"]. An unpublished diploma thesis. Olomouc: Univerzita Palackého.
Wales, K. (2001) *A Dictionary of Stylistics.* Harlow: Pearson Education Limited.

Sources

Crafts, N. (2004) 'Steam as a general purpose technology: A growth accounting perspective.' *The Economic Journal, 114*, 338-351. [Text 2]
Parker, S., Belghitar, Y. and Barmby, T. (2005) 'Wage uncertainty and the labour supply of self-employed workers.' *The Economic Journal, 115*, 190-207. [Text 1]
Winters, L. A. (2004) 'Trade liberalisation and economic performance: An overview.' *The Economic Journal, 114,* 4-21.