

## A Tightrope with a Net

Claudia Monacelli

Volume 49, Number 4, December 2004

URI: <https://id.erudit.org/iderudit/009788ar>

DOI: <https://doi.org/10.7202/009788ar>

[See table of contents](#)

### Publisher(s)

Les Presses de l'Université de Montréal

### ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

### Cite this article

Monacelli, C. (2004). A Tightrope with a Net. *Meta*, 49(4), 891–909.  
<https://doi.org/10.7202/009788ar>

### Article abstract

This is a small-scale, corpus-based study on the effects of lexical repetition, as a cohesive device, on textual coherence. Following Hoey's work on patterns of lexis (1991), we introduce categories of repetition to consider for 'automatic' processing using a concordancer program in order to produce summaries based on clusters of lexical repetition sequences, thus obtaining an outline of a text's structure. The study's findings are extended to the analysis of parallel texts: two professional translations of corpus text 1 are examined using the same procedure.

# A Tightrope with a Net

CLAUDIA MONACELLI

University of Bologna, Forlì, Italy  
claudiamonacelli@libero.it

## RÉSUMÉ

Basée sur un corpus limité, cette étude examine les effets de la répétition lexicale, comme instrument de cohésion, sur la cohérence textuelle. Suivant le modèle de cohésion lexicale de Hoey (1991), on utilise ses catégories de répétition qui s'adaptent à l'élaboration 'automatique' avec un concordancier pour en extraire des sommaires basées sur des groupes de séquences lexicales répétées, et obtenir ainsi la délimitation structurelle du texte. Les résultats sont étendus à l'analyse de textes parallèles: deux traductions professionnelles du texte n° 1 du corpus sont examinées suivant la même procédure.

## ABSTRACT

This is a small-scale, corpus-based study on the effects of lexical repetition, as a cohesive device, on textual coherence. Following Hoey's work on patterns of lexis (1991), we introduce categories of repetition to consider for 'automatic' processing using a concordancer program in order to produce summaries based on clusters of lexical repetition sequences, thus obtaining an outline of a text's structure. The study's findings are extended to the analysis of parallel texts: two professional translations of corpus text 1 are examined using the same procedure.

## MOTS-CLÉS/KEYWORDS

lexis, repetition, intertextuality, translation, concordancer

## 1. Introduction

In his book entitled *Patterns of Lexis* (1991), Michael Hoey investigates the text-organizing function of lexical repetition. He argues it is possible to identify marginal and central sentences in a passage through a systematic analysis of lexis in text. By eliminating marginal sentences and combining central ones, he claims to develop "a methodology for the production of readable abridgements of texts that is capable of some degree of automation" (Hoey 1991:3). In terms of translation, summarizing a text requires us to pay attention to its structure, which then becomes "an important guide to decisions regarding what should or should not appear in the derived text" (Hatim and Mason 1990:185). The present study has three aims: to explore the second part of Hoey's claim, namely the degree of automation possible in the production of text abridgements using his model; to analyze the role of lexical cohesion in fostering textual coherence; to consider the possibility of using Hoey's model to assess parallel texts in translation studies.<sup>1</sup>

The study of Hoey's work stems from interest in adapting his model for use in the analysis of lexical patterns in oral texts. His work, however, strictly involves written texts, based on the analysis of complex lexical patterns running across sentences to form nets. And this, of course, represents the first problem in even attempting to adapt such a model. Indeed Hoey's use of the sentence as a unit inspired the title of

this essay. A sentence – like a tightrope – has a crisply defined beginning and end. The sentence as a unit of study is similar to walking a tightrope: even though it is common practice to write in sentences, people don't talk in sentences, they don't think in sentences, nor do they usually walk tightropes. The implication here is clearly one of risk; unless there is a net. And this is where Hoey's model becomes of interest.<sup>2</sup>

The corpus examined consists of 10 texts extracted from academic textbooks in English that are classified, according to the Dewey Decimal Classification System, as belonging to the social sciences (300). We analyze how an introduction develops the textbook's themes once it is abridged, that is, once it has been processed using those features of Hoey's lexical repetition model that lend themselves to automatic processing. Two professional translations of corpus text 1 are also examined using Hoey's model of lexical repetition.

We begin by introducing Hoey's views on lexical cohesion and textual coherence (§2), and examine a text's 'organization' and 'structure' (§3). Hoey's taxonomy of lexical repetition categories considered for 'automatic' processing is presented using examples from corpus text 1 (table 1). Results are examined (§5, §6) and findings are then extended to possible applications in the assessment of parallel texts (§7). A general discussion weighs the values and pitfalls of this approach (§8) and we offer suggestions for the development of future work (§9).

## 2. Lexical cohesion and textual coherence

How cohesion works and how it serves to explain what happens in a text depends largely on the literature espoused for these descriptions. Perhaps the most widely read work on cohesion is Halliday and Hasan's *Cohesion in English* (1976). For them, 'texture' (organization) consists of relations among items in a text which they call 'cohesive ties' (Halliday & Hasan, 1976:2-3), and they establish five classes of cohesion: conjunction, reference, substitution, ellipsis, and lexical cohesion (*ibid.*:29). The authors further distinguish two subclasses under lexical cohesion: reiteration and collocation. Michael Hoey specifies that – with the exception of conjunction – these classes of cohesive ties are all ways of repeating (Hoey, 1991:6). He is quick to point to the fuzzy boundaries between different forms of reiteration and stresses how, nonetheless, all types of lexical reiteration, along with collocations, establish *lexical* relations and only in a second instance do they mark textual ones (Hoey, 1991:6-7).

As he states his case for the importance of lexical cohesion, Hoey compares the frequency of the different ties Halliday and Hasan find in their own analyses and indicates that lexical cohesion alone accounts for nearly 50% of all ties. He mentions, however, that lexical items may form a relationship with more than one other item, thus showing how Halliday and Hasan's data has not accorded lexical cohesion the importance it deserves. He concludes that "lexical cohesion is the only type of cohesion that regularly forms multiple relationships (though occasionally reference does so too). If this is taken into account, lexical cohesion becomes the dominant mode of creating texture" (Hoey 1991:9).

When discussing cohesion as a device which establishes relations between grammatical or lexical items in a text, the question of textual coherence inevitably surfaces.

As Hoey points out, cohesion is not synonymous with coherence for many scholars (cf. Widdowson, 1978; Beaugrande & Dressler, 1981). He himself assumes cohesion to be a property of the text and coherence to concern the reader's evaluation of a text and suggests looking at text patterning, created by lexical cohesion, which relates to the ways in which topics interrelate in their development. Hoey uses the notion of 'topic shift,' signaled by lexical clusters forming bonds, as an *indication* of text structure. In fact, he treats the issue of structure in text as open for debate and concedes, however, that "there still may be structure in a looser or different sense" (*ibid.*: 29).

Echoing a systems theory perspective and a view at the basis of practically all work on discourse,<sup>3</sup> Hoey reminds the reader that sentences have a meaning together that is more than the sum of their separate parts. In essence, this is the view characterizing the paradigm shift throughout the 20th century in other branches of science, a process-oriented view.

### 3. Text organization and discourse structure

Systems thinking was pioneered by biologists who emphasized the view of living organisms as integrated wholes. The basic tension is one between the parts and the whole: the essential properties of an organism or living system are properties of the whole, which is more than the sum of its parts. The paradigm shift involves contextual thinking, putting phenomena into the context of a larger whole. An emphasis on process thinking began making its way into several realms: beginning with von Bertalanffy in the '30s, who defined as 'open systems' any living structure that depended on flows of energy and resources,<sup>4</sup> and continuing with the cybernetic movement of the '40s which introduced the concepts of feedback loops and dynamic systems (Capra, 1997:58-64). But it wasn't until the '70s that Ilya Prigogine used the term 'dissipative structures' to describe the new thermodynamics of open systems as combining the stability of structure with the fluidity of change (*op. cit.*:180).

Around the same period, the beginning of the '70s, Maturana and Varela advanced their theory of autopoiesis, which essentially views living organisms as operationally closed entities which subordinate all changes to the maintenance of their own organization. Living organisms have a distinct structure which is continuously recreated through interactive feedback cycles (Maturana & Varela, 1980; 1998). In his seminal work entitled *The Web of Life* (1997), Fritjof Capra brings together the concepts of autopoiesis and dissipative structures and defines the link as one between organization and structure, cognition – as process – being the inextricable link.

Dissipative structures call to mind the notion of intertextuality which problematizes the idea of a text as having boundaries, the boundaries of texts being permeable. The semiotic concept of intertextuality, as introduced by Kristeva, refers to texts in terms of two axes: a *horizontal axis* connecting the author and reader of a text, and a *vertical axis*, which connects the text to other texts (Kristeva, 1980:69). She argues that rather than confining our attention to the structure of a text we should study its 'structuration' (how the structure comes into being, or its construction). Intertextuality here is considered access to texts via our knowledge of encountered text "in a continual process of reconstruction of our individual and social realities (Seidlhofer, 2000:211).

Hatim and Mason distinguish intertextuality as "an ideal testing ground for basic semiotic notions in practical pursuits such as translating and interpreting"

(1990:121). They illustrate how intertextual chains permeate texts, forming “strands of reference to previous knowledge enshrined in texts we have encountered” (1990:123) which makes it truly a dynamic property of texts. They mention ‘active’ and ‘passive’ intertextual links: ‘active’ links activate knowledge and belief systems beyond the text itself; passive links aim to maintain a text’s internal coherence (1990:123-124). Examples of these forms of intertextuality are taken from corpus text 1 (sample 3) which has a total of 17 sentences. Sentences are numbered for convenience of reference and the text’s original division into paragraphs has been maintained.

### *Sample 3 Corpus text 1*

1 This book is a textbook based on original research and develops an important thesis. 2 It concerns the globalization of social policy and the socialization of global politics. 3 The book demonstrates first that national social policy is increasingly determined by global economic competition and by the social policy of international organizations such as the World Bank, and secondly that the substance of social policy is increasingly transnational. 4 Global social policy is constituted of global social redistribution, global social regulation and global social provision and empowerment. 5 This textbook reviews for students the state of the world’s welfare in terms of how far human needs are met. 6 Trends in global inequity, and diverse experiences of different kinds of welfare regime North, South, East and West, are summarized. 7 The social policies of international organizations are reviewed systematically for the first time. 8 The book is also a report of two major research projects which focused on the making of post-communist social policy and the role played in this by international organizations in Bulgaria, Hungary, Ukraine and the post-Yugoslav countries. 9 The research which documents the global discourse taking place within and between international organizations about the future for welfare policy reinforces the thesis concerning the globalization of social policy.

10 The book is primarily addressed to students of social policy but it is also intended that it should be read by other colleagues and students in academia, government and international organizations. 11 Development studies specialists would benefit from the comparisons and connections made between social policy in developing and developed countries. 12 Students of international organizations and international relations will find the book informs the debates concerning the future of these areas of study. 13 Economists should read it because it demonstrates that choices between economic and social policies are a matter not only of mathematical modelling but of political values. 14 Political scientists should read it because it demonstrates that the locus of key political decisions lies far from national governments and inside global banking organizations. 15 Soviet and East European area studies specialists should read it to appreciate the importance of external influences on the region. 16 Sociologists should read it as an example of an attempt to study the social relations of power and the nature of discursive practices at a global level. 17 Social policy makers and their advisers should read it to locate the institutions within which it might be most appropriate to apply their skills.

A cursory reading of this non-narrative text brings to light several characteristics: a parallel structure in sentence 2 with a transposition of parts of speech (globalization/

social, socialization/global); a fair amount of lexical repetition, e.g. 'global' repeated no less than four times in sentence 4 and 'should read it' in sentences 13 to 17. Both examples cited involving repetition are cases of passive intertextuality. Active intertextual strands within the text involve the 'World Bank' (international organizations, mathematical modelling, banking organizations, etc.), 'post-communist social policy' (Bulgaria, Hungary, etc.).

#### 4. Methodology and corpus

Table 1 lists Hoey's categories of lexical repetition which best lend themselves to processing using a concordancer. Examples refer to corpus text 1 (sample 3). We have not considered the categories of simple partial paraphrase (after Hoey), substitution, co-reference, and ellipsis, which require a great deal more 'manual' processing/analysis.

TABLE 1

##### Lexical repetition categories considered for automatization

Category	examples
<i>simple repetition</i>	<b>social</b> (sentences 2 and 3); <b>government</b> (sentence 10), <b>governments</b> (sentence 14)
<i>complex repetition</i>	<b>politics</b> (sentence 2), <b>political</b> (sentence 13); <b>important</b> (sentence 1), <b>importance</b> (sentence 15)
<i>simple mutual paraphrase</i>	<b>book</b> (sentence 1), <b>textbook</b> (sentence 5)
<i>antonymous complex paraphrase</i>	<b>national</b> (sentence 3), <b>international</b> (sentence 7)
<i>other complex paraphrase</i>	link triangle: <b>national</b> (sentence 3), <b>international</b> (sentence 7), <b>global</b> (sentence 2)

There are many shades of grey in the establishment of simple and complex repetition, as Hoey himself points out, and he offers a series of decision-making flow charts (*ibid.*:58-60) as a guide in seeking out phenomena which are worth investigating in a text. He explains how sentences come to be viewed as being *central* or *marginal* by the number of lexical links they have with other sentences. The criterion used in establishing links in a long text may be difficult to control, but Hoey adds an important constraint,

if a lexical item appears for the third (or more) time in a text, it is only necessary to establish a contextual connection with *one* of the previous occurrences for the item to be treated as forming a repetition link with *all* the previous occurrences. (*ibid.*:57, original emphasis)

The corpus examined,<sup>5</sup> was compiled by randomly selecting textbooks in English off the Ruffilli Library<sup>6</sup> shelves in the social sciences section. The texts are introductions to the volumes they precede and, as such, it is safe to say that readers would expect these texts to fulfill an introductory function and look to them for guidance in terms of what lies ahead in the textbook, following the principle of *relevance*.<sup>7</sup> These introductions appear in the volumes within the context of other pages in the book which include information such as: title, author/s, editor/s, publishing house, date, notes on

copyright, table of contents, dedication, acknowledgements. The reader is thus in a position of (possibly) knowing what the book is about or, in any case, it is presumed that s/he reads an introduction for the purpose of confirming (or disconfirming) any expectations s/he may have in relation to the textbook, and/or that the introduction serves to create further expectations concerning the entire volume.

The corpus texts come under different names in each volume. Three texts are called 'introduction' (texts 2, 5, 7), four are a 'preface' (texts 1, 3, 6, 8, 10), and two are labelled 'foreword' (texts 4, 9). Since they were chosen as representative of short introductions to textbooks, which inform readers of the book's contents, we excluded any acknowledgements, if they were included as part of the text.

#### 4.1 Preparation of texts

Texts were transcribed and saved in '.txt' format. This removed their original graphical layout (paragraph divisions, indentation, skipped lines, etc.). Sentences were marked with a full stop, followed by a space and a capital letter, e.g. ' M'. The concordancer *Concordance* is programmed,<sup>8</sup> by default, to read lines as they appear on a '.txt' file, hence our text files consist of a series of single sentences, one below the other, with no spaces between them, nor any indentation. Indented quotes in texts are treated as part of the preceding sentence, and include the citation (e.g. '(Gramsci in Booth, 1991:1)'), corpus text 5). Authors use citations to make an overt, intertextual link for the reader. In terms of cohesion they are examples of how an author attempts to use intertextuality to enhance coherence for the reader since, as discussed, intertextual links are necessary for a text to be cognitively perceived as coherent. The program also disregards any punctuation unless otherwise specified, which means it considers as single lexical items those which are hyphenated (e.g. 'self-access' or 'intra-textual') and any compounded forms with a slash (e.g. 'type/token').

Footnotes are excluded from our study as a matter of convenience. It might be argued that they function in a similar manner as quotations, offering overt intertextual links. However, since authors themselves excluded end-notes and footnotes from the text proper, we too have disregarded them, as they – in no case studied – informed the reader of the book's contents.

Criticism may be leveled as to the 'flattening' of the graphical layout in our texts in terms of what it potentially communicates to a reader. For example, writers – it is supposed – divide their texts into paragraphs for a reason, thus giving 'shape' to their text, for what it is worth. This 'shape,' in turn, may influence the text receiver into compartmentalizing the words-on-the-page into well-defined groups, as suggested by the graphical layout. But since our study concerns the power of lexical repetition as a cohesive device, and its role in fostering coherence for the reader, whether we skip a line, indent, highlight, underline or italicize is of little consequence. In fact, we have found that sentences appearing in the abridged form of an introduction have little, or nothing, to do with their original position in the author's text. In other words, the first sentence of a paragraph may not necessarily be one which introduces new concepts or a topic shift, in relation to the context, nor does it necessarily mean that, because of its 'position,' it should be considered a 'central' sentence, to use Hoey's term.<sup>9</sup>

#### 4.2 A net of links

Table 2 illustrates data relating to the ten corpus texts: the number of word ‘types,’ word ‘tokens’ (the running words in each text), the type/token ‘ratio,’ and the average number of words in a sentence (‘w/sent.’). Many authors warn that the type/token ratio is a poor indicator of lexical density, and this also in terms of its application to different languages. However, we are interested in the type/token ratio and the average lengths of sentences for what they can tell us about a cut-off point for sentence pair bonding.

TABLE 2

##### Lexical density in corpus texts

<i>corpus</i>	1	2	3	4	5	6	7	8	9	10
types	179	246	258	299	367	248	294	338	387	311
tokens	397	485	525	759	1028	467	710	662	990	690
ratio	2.22	1.97	2.03	2.54	2.80	1.88	2.42	1.96	2.56	2.23
w/sent.	23.35	26.94	27.63	29.19	28.56	17.96	20.29	20.06	30.94	34.65

Repetition, by (commonly-accepted) definition, is something that occurs *after* a phenomenon, which it replicates in some way. In this sense, repetition is typically assumed to be an anaphoric link. However, since Hoey is interested in repetition that appears to serve some text-organizing function, he uses the statistical data of links between sentences in a text to establish significant connections between sentences. Lexical items form links between sentences and sentences with three or more links are considered to form bonds. A bond is not defined in absolute terms; a cut-off may be adjusted to suit a text, especially if a high percentage of sentences share three links. The cut-off point is related, to some extent, to the length and text density of sentences of a text: the higher the text density, the higher the cut-off point (*ibid.*:90-91).

#### 4.3 A repetition matrix of corpus text 1

Table 3 is a matrix with the sentence numbers of our text as both its parameters. It is used to record the number of links identified between items across sentence boundaries. Rows tell us whether a sentence connects to previous ones; columns tell us whether sentences connect to subsequent ones. For example, sentence 10 (row 10) shares links with all previous sentences. Column 10 indicates that it also shares links with all subsequent sentences. The matrix also reveals a density of connections which varies throughout the text. We can see how some sentences appear to be closely linked to other sentences of the text (e.g. sentences 3, 7, 10), whereas others only slightly connect or do not at all (e.g. sentences 1, 13).



TABLE 3  
**Repetition matrix of corpus text 1**

1																	
2		2															
3	1	3	3														
4		3	3	4													
5	1	1	2	1	5												
6		2	2	2	2	6											
7		3	<b>5</b>	3	2	2	7										
8	3	3	<b>5</b>	3	2	2	4	8									
9	2	4	4	3	2	3	4	4	9								
10	2	3	<b>5</b>	3	3	2	4	<b>5</b>	4	10							
11	2	2	3	2	2	1	2	4	3	4	11						
12		2	2	1	2	1	1	2	4	2	1	12					
13		2	4	2		1	2	2	3	4	3		13				
14		2	<b>5</b>	2	2	1	2	3	3	<b>6</b>	2	2	4	14			
15	1	1	1	1	2	1	1	2	1	4	2	3	1	2	15		
16		2	2	2	2	2	2	2	3	<b>5</b>	2	3	3	3	3	16	
17		2	2	2		1	2	3	2	4	2		4	4	1	3	17

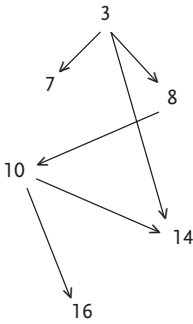
In order to consider repetition as a text-organizing function, Hoey suggests concentrating on those cases that show “an above-average degree of connection” (Hoey, 1991:91). The choice of a cut-off for sentence bonding was taken in relation to the original text length and to the proportion of sentence pairs that would be included for any given cut-off point. For example, the matrix in tab. 3 shows a total of 51 sentence pairs sharing bonds formed by three or more links. A cut-off of three links for corpus text 1 would yield a total of 17 sentences and, of course, this simply tells us the text is cohesive to a certain degree and would not help us abridge it. We opted for a cut-off of five links, amounting to 6 sentences in the abridged version. We now extract the relevant information and analyze the ‘net’ of bonds created, to see what it can tell us about how text 1 is organized.

#### 4.4 Creation of a net of bonds

Hoey makes a point of using nets to illustrate ‘topic-opening’ and ‘topic-closing’ sentences, but also to show where subtopics emerge. Using the information in the matrix as a point of reference (tab. 3), we now attempt to illustrate how the sentence pairs that form a bond interconnect. Sentence 3 bonds with sentences 7, 8 and 14; sentence 8 bonds with sentence 10; sentence 10 bonds with sentences 14 and 16 (table 3). Figure 3 (strictly a topological diagram) shows each pair of bonded sentences as represented by a line. The relation between sentence order and the vertical axis of the diagram is maintained, even though the length, position or angle of lines have no importance.

FIGURE 1

Net of bonded pairs found in corpus text 1, tab. 3



Sentence 3 in the diagram seems to be a ‘topic’ sentence, introducing possible subtopics or developments of the main topic. A subtopic may be introduced by sentences 8 and 10, which are further developed or clarified by sentences 14 and 16.

We now have a record of the number of bonds each sentence shares, both with previous and subsequent sentences. Since bonding was found to be unevenly distributed in our texts, we now examine how these bonds function. Below is a list of all sentences in corpus text 1 and in parentheses next to the sentence number are two co-ordinates: the first co-ordinate refers to the number of bonds the sentence shares with previous sentences; the second co-ordinate refers to the number of bonds it shares with subsequent sentences. Hoey explains the variation in the distribution of bonding by illustrating that sentences with a high second co-ordinate are *topic-opening* and those with a high first co-ordinate are *topic closing* (Hoey, 1991:119).

TABLE 4

Topic opening and topic closing sentences: text 1

1	(-,0)	10	(2,2)
2	(0,0)	11	(0,0)
3	(0,4)	12	(0,0)
4	(0,0)	13	(0,0)
5	(0,0)	14	(2,0)
6	(0,0)	15	(0,0)
7	(1,0)	16	(1,0)
8	(1,1)	17	(0,0)
9	(0,0)		

Sentence 3 shares bonds with four subsequent sentences (tab. 3), thus acting as a topic opener. Sentence 7 shares one link with a previous sentence (sentence 3), while sentence 8 seems to have a ‘central’ role, since it shares bonds with both a previous and a subsequent sentence. Sentence 10 behaves in much the same way, sharing two bonds with sentences before and after it. On the other hand, sentences 14 and 16 are topic closing sentences, only sharing bonds with previous sentences.

The summary of corpus text 1 is presented as sample 6 below. The removal of marginal sentences has made for a rather smooth-reading summary. A closer analysis of sentences removed shows how they express concepts included in some form in the

summary's 'central' sentences: such as sentence 1 ('This book is a textbook based on original research and develops an important thesis.') included in sentence 8; sentence 13 ('Economists should read it because it demonstrates that choices between economic and social policies are a matter not only of mathematical modelling but of political values.') included in sentence 14; sentence 15 ('Soviet and East European area studies specialists should read it to appreciate the importance of external influences on the region.') included – or alluded to – in sentence 8.

### *Sample 6 Summary of corpus text 1*

3 The book demonstrates first that national social policy is increasingly determined by global economic competition and by the social policy of international organizations such as the World Bank, and secondly that the substance of social policy is increasingly transnational. 7 The social policies of international organizations are reviewed systematically for the first time. 8 The book is also a report of two major research projects which focused on the making of post-communist social policy and the role played in this by international organizations in Bulgaria, Hungary, Ukraine and the post-Yugoslav countries. 10 The book is primarily addressed to students of social policy but it is also intended that it should be read by other colleagues and students in academia, government and international organizations. 14 Political scientists should read it because it demonstrates that the locus of key political decisions lies far from national governments and inside global banking organizations. 16 Sociologists should read it as an example of an attempt to study the social relations of power and the nature of discursive practices at a global level.

The book's aim is clearly stated in sentence 3; sentence 7 enhances the notion that the authors contribute to the scientific community ('are reviewed systematically for the first time'), thus establishing credibility; specific research projects are mentioned in sentence 8, even if these are set in the 'background' ('this book is *also* a report'); primary intended readers are mentioned in sentence 10, which also suggests the book 'should be read' by a series of others; sentences 14 and 16 specifically name further intended readers, in both cases particular members of the scientific community. Generally speaking, the emergent discourse is one where authors lay claim to authority in order to convince the readers to buy their argument.

The following section examines the results of applying Hoey's lexical repetition model to our corpus. The findings are then extended to the analysis of two translations of corpus text 1 (§7).

## **5. Results**

We now consider the resulting abridged versions by comparing the data in tab. 5 to see if we can say anything about how lexical density and sentence length influence the process of abridgement.

Texts 2 and 6, which are of similar length (484 and 476 words, respectively) and have a similar type-token ratio (1.97 and 1.88 respectively), are worth comparing. Since the lexical density of both texts is similar, why is it that in one text – text 2 – a cut-off of five links was set for the establishment of a bond, and in the other – text 6 – a cut-off of three links was used? The answer lies in the length of sentences in both

texts: text 2 has an average of almost 27 words per sentence, whereas in text 6 the average number of words per sentence is about 18.

TABLE 5

**Lexical density in relation to text abridgements**

<i>corpus</i>	1	2	3	4	5	6	7	8	9	10
types	179	246	258	299	367	248	294	338	387	311
tokens	397	485	525	759	1028	467	710	662	990	693
ratio	2.22	1.97	2.03	2.54	2.80	1.88	2.42	1.96	2.56	2.23
w/sent.	23.35	26.94	27.63	29.19	28.56	17.96	20.29	20.06	30.94	34.65
no.sent.	17	18	20	25	36	26	35	32	31	19
cut-off	5	5	7	7	6	3	4	4	5	6
abridged	6	4	4	6	6	11	11	9	11	6

Text 5 and 7 are also similar in terms of the number of sentences in the original versions: 36 in text 5 and 35 in text 7. Here, too, there is diversity in the number of sentences which survived the process of abridgement: 6 in text 5 and 11 in text 7. This difference is also a result of the average number of words per sentence: 29 in text 5 and 20 in text 7.

Although texts 2 and 3 are similar in terms of certain data categories (type/token ratio, average words per sentence), they differ greatly as to the number of links used as a cut-off for bonding (5 in text 2 and 7 in text 3), but both have the same number of sentences in their abridged versions (4). This indeed comes as a surprise. However, although similar, text 3 is in fact longer: 525 running words compared to 485 words in text 2. In this case the higher cut-off used in text 3, seven links, served to reduce the text even further. Had we adopted a cut-off of six links for text 3, we would have had 11 sentences in the abridgement which, however, is merely half the original text's length.

It seems the longer the sentence length, the higher the cut-off set for sentence bonding, in relation to the type-token ratio. We bear these points in mind as we turn to analyse two translated versions of corpus text 1.

## 6. The assessment of parallel texts in translation

Table 6 lists information concerning both corpus text 1 and two translations of this text. We notice that the Italian versions are rather longer than the original text. It also seems the two translators approached the text from two different perspectives: translator 1 seems to have maintained the original number of sentences, thus mirroring the source text to some extent; translator 2, on the other hand, reduced the number of sentences in relation to the source text, but increased the number of words per sentence. This strategy influenced the cut-off point in establishing inter-sentential bonds: 8 in translation no. 2, compared to 5 in the source text. There is also a higher cut-off point in translation no. 1, due obviously to the longer sentences in the translated version: 6 compared to 5 in the source text.

TABLE 6

**Lexical density in relation to text abridgements: corpus text 1 and translations**

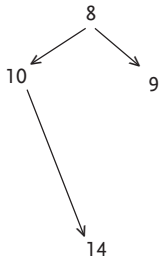
<i>text</i>	1	Tran 1	Tran 2
types	179	236	261
tokens	397	491	520
ratio	2.22	2.08	1.99
w/sent.	23.35	28.88	37.14
no.sent.	17	17	14
cut-off	5	6	8
abridged	6	4	4

The following sections review the resulting summaries and a final discussion (§8) weighs the pros and cons of using Hoey's model to assess translations.

*6.1. Translation no. 1*

Figure 4 illustrates the resulting net of bonded pairs in the first translation of corpus text 1. Compared to the source text net (fig. 3) which 'opened' with sentence 3 as a topic sentence and was articulated through sentences 8 and 10, this translated version foregrounds sentence 8 as topic opener and sentence 10 is also a topic or subtopic sentence.

FIGURE 2

**Net of bonded pairs found in first translation of corpus text 1**

The summary (sample 7) begins with a sentence which includes 'unwanted' cohesion in sentence 8 ('*anche*,' also). Hoey removes all local, non-lexical cohesive features, since they interfere with the connections being made with other sentences in the text, once they are abridged. These are indicated as follows: <also>. A gloss of the Italian text is included in square brackets.

*Sample 7 Summary of corpus text 1: translation 1*

8 Il volume fa <anche> riferimento a due importanti progetti di ricerca che si sono occupati di come sia stata realizzata la politica sociale nei paesi post-comunisti e del ruolo svolto da organizzazioni internazionali in Bulgaria, Ungheria, Ucraina e nei paesi della ex-Yugoslavia. 9 Queste ricerche, che documentano il discorso globale che

si svolge all'interno delle organizzazioni internazionali e fra le medesime riguardo al futuro delle politiche di welfare, rafforza la tesi che riguarda la globalizzazione della politica sociale. 10 Questo volume è rivolto principalmente a studenti di politica sociale ma può essere utile anche a colleghi e studenti di altri settori accademici e a coloro che operano in istituti governativi e organizzazioni internazionali. 14 Gli studiosi di scienze politiche troveranno il volume di loro interesse perché viene dimostrato che la sede delle decisioni politiche fondamentali si trova lontano dai governi nazionali ed è all'interno di organizzazioni finanziarie globali.

[8 The volume <also> makes reference to two important research projects which have dealt with how social policy in post-communist countries was accomplished and with the role played by international organizations in Bulgaria, Hungary, the Ukraine and former Yugoslavian countries. 9 These research projects, which document the global discourse which occurs within international organizations and among the same concerning the future of welfare policies, strengthens the thesis that concerns the globalization of social policy. 10 This volume is mainly aimed to students of social policy but can also be useful to colleagues and students of other academic sectors and to those who operate in government institutes and international organizations. 14 Scholars of political science will find the volume of interest because it is shown that the place where fundamental political decisions are taken is far from national governments and is within global financial organizations.]

In the summary of translation no. 1 (sample 6), we see how the translator foregrounds the two research projects (sentence 8) and the fact that the text is intended for students of social policy and other students and colleagues in academia (sentence 10). The source text structure (fig. 3), on the other hand, foregrounds the *result* of the two research projects ('the book demonstrates... firstly... and secondly,' sentence 3), then introduces the research projects ('The book is *also* a report,' sentence 8). Sentence 10 in this version mentions intended readers and others, while sentence 14 – the only sentence of those listing people who 'should read' the book which survived the summary process – suggests the book for '*scholars* of political science,' whose meaning potential only extends to an academic context.

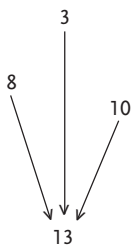
Differently from the source text summary, this abridged version excludes any reference to 'sociologists' (sentence 16) or 'social policy maker's' (sentence 17), as benefiting from the volume.

## 6.2. Translation no. 2

Figure 3 illustrates the resulting net of bonded pairs in the second translation of corpus text 1. This summary opens with sentence 3 as a topic sentence, similar to the source text summary (sample 6). Here, too, we find sentences 8 and 10 as signalling topic shifts. However, these three all 'point' to sentence 13.

FIGURE 3

## Net of bonded pairs found in second translation of corpus text 1



The complete second translation of corpus text 1 includes a total of 14 sentences: sentences 13, 14, 15 and 16 of the source text are included in sentence 13 of this translation, through the use of semi-colons to separate concepts. This strategy may be considered one of compensation, since the effect of the repeated lexical items in the source text ('should read it') in conveying insistence that the book is intended for a series of other (just as important) readers, is indeed here conveyed by turning the recurrent verb phrase into recurrent prepositions ('advised *to* economists; ... *to* political experts; ... *to* specialists of; ... *to* sociologists').

*Sample 7 Summary of corpus text 1: translation 2*

3 Con questo libro si vuole dimostrare, da un lato, che le politiche sociali nazionali sono determinate, in maniera sempre più evidente, dalla competizione economica globale e dalle politiche sociali attuate da organismi internazionali (quali ad esempio la Banca Mondiale), e dall'altro, che l'essenza delle politiche sociali è sempre più *trans*-nazionale. 8 Il testo può anche essere considerato come resoconto dei due principali progetti di ricerca che hanno focalizzato la propria attenzione sulla formazione delle politiche sociali nelle società post-comuniste e sul ruolo giocato in questo senso dagli organismi internazionali in paesi quali Bulgaria, Ungheria, Ucraina, ed ex-Yugoslavia. 10 Questo libro è innanzitutto indirizzato a studenti di Politiche Sociali, ma è sottinteso che possa anche essere letto da colleghi e studenti all'interno delle università, della pubblica amministrazione, e degli stessi organismi internazionali. 13 La lettura di questo libro è, inoltre, consigliata agli economisti poiché esso dimostra che le scelte tra politiche economiche e sociali non sono una semplice questione di modelli matematici, ma implicano decisioni politiche; ai tecnici della Politica poiché il nostro testo dimostra che le sedi deputate alle decisioni politiche chiave non sono i governi nazionali bensì gli organismi bancari internazionali; agli specialisti di questioni Sovietiche e dell'Est Europa affinché possano comprendere appieno l'importanza delle influenze esterne su questa regione del mondo; ai sociologi che potranno interpretare il libro quale tentativo di studio dei rapporti fra la sfera sociale e quella delle decisioni politiche e delle forme di partecipazione alle decisioni di livello internazionale.

[3 With this book we wish to show, on the one hand, that national social policies are determined, in an ever more evident fashion, by global economic competition and by social policies implemented by international organizations (such as, for example,

the World Bank), and on the other, that the essence of social policies is ever more transnational. 8 The textbook may also be considered a report of two main research projects which have focussed their attention on the formation of social policies in post-communist societies and on the role played in this sense by international organizations in countries such as Bulgaria, Hungary, the Ukraine, and former Yugoslavia. 10 This book is above all addressed to students of political science, but it is understood that it may also be read by colleagues and students within universities, the public administration, and the same international organizations. 13 The reading of this book is, moreover, advised to economists since it shows that the choices between economic and social policies are not simply a question of mathematical models, but imply political decisions; to political experts since our text shows that the places where key political decisions are made are not national governments but international banking institutes; to specialists of Soviet and East European issues so that they might fully understand the importance of external influences on this region of the world; to sociologists who could interpret the book as an attempt to study the relations between the social sphere and that of political decisions and of the forms of participation in international decision-making processes.]

It is interesting to see how the two research projects are now relegated even farther into the background in this translation by the addition of a modal ('the textbook *may also* be considered,' sentence 8). Sentences 10, which lists the primary intended readers, is enhanced by sentence 13, listing yet other intended readers with marked emphasis, as if it were through these that we understand the author's (or the translator's) underlying intention. Here, the translator refers to political scientists as 'tecnicisti della Politica' ['technicians of Politics,' or political experts], whose meaning potential only extends to a professional context ('doing' political science). And this is in stark contrast to the 'scholars' in translation no. 1. All in all the translator seems to convey an underlying tension between the 'local' (the textbook intended for students of social policy, national social policy, etc.) and the 'global' (other students and colleagues, economists, political scientists, global banking organizations, etc.), not unlike the author of the source text.

These discursual activities emerge after having processed the texts (both source texts and translations) using Hoey's model and a concordancer. Our analysis of the two translations is far from exhaustive and simply aims to assess the value of this method applied to the study of parallel texts. Are we satisfied with what this process is able to tell us? Does it tell us anything we wouldn't have been able to know by simply reading them? In other words, does it inform us on the 'quality' of these translations?

## 7. Discussion

In §4 we mentioned Hatim and Mason's notion of intertextual chains and 'active' and 'passive' intertextuality. Having now examined Hoey's lexical repetition model we can say, for example, that 'passive' intertextuality is comparable to Hoey's categories of lexical repetition considered for automatic processing (tab. 1). Hatim and Mason stress that coping with passive forms of intertextuality is vital for a translator, since "reiteration of text items is always motivated" (1990:124). But what about the



more 'active' chains of intertextuality, intertextuality which signals entire discursive structures? Items such as 'World Bank' (corpus text 1, sample 3) are composed of two items and Hoey's model only contemplates the lemmatization (grouping together related words under a single headword) of single lexical items.

In an interesting study on the use of (an adapted version of) Hoey's model to assess translations, Klaudy and Károly (in Olohan 2000:143-157) theoretically resolve the "particular semantic and structural problems" in Hoey's taxonomy (cf. Károly, 1999). Their 'lexical items' also extend to include phrasal compounds often used together to refer to a unique concept. They separate lexical repetition categories into 'lexical relations' and 'text-bound relations,' the latter covering what they call 'instantial relations,' which are further subdivided into relations of 'equivalence' and 'naming.' Klaudy and Károly used this revised model to seek information about the global meaning of a text, based on linguistic elements detected on the textual surface. More precisely, their analysis focused on distinguishing the text-building strategies of professional translators from those of trainee translators. Although their study revealed differences in the two groups' use of lexical repetition as a cohesive device, they also found indications of differences in the discourse strategies applied (*ibid.*:143). The authors identified the 'structural' components of the text (argumentative text structure), counted repetition links and bonds within the text, and statistically compared the two groups' use of repetition in their target texts, and in relation to the source text. They concluded that professional translators use repetition and bonding in a way more similar to the source text than novice translators (*ibid.*:152) and that their analytical tool was capable of highlighting differences in the quality of translations. In expressing caution about making deductions of the type Klaudy and Károly make in their study, we quote Mason on cohesion and coherence in source (ST) and target texts (TT):

It is also of course the case that each text – ST and TT – is cohesive in its own terms. The expectation that ST cohesive use is necessarily, or even desirably, transferable to a TT is, in itself, a naïve one, stemming from a view of translating as language transfer rather than as motivated behaviour within a particular context and responding to its own norms. (Mason, 2001:73)

Mason's words remind us of those notions propagated by descriptions of cohesive devices in the literature which we cannot whole-heartedly embrace. Hoey's personable writing style and the common sense quality of his statements truly seduce the reader. There are, however, moments when the words-on-the-page literally jump out at you because of their incongruence with possibly held, deep-rooted philosophical positions. Such is the case (at least for *this* reader) when he insists on imbibing a text with 'information content,' as if it were possible to truly measure its consistency. But regardless of the epistemological perspective with which a reader tints Hoey's work, it is undeniable that the model he advances is 'filled' with a fascination all its own: discourse structures seem to emerge from a net of bonded sentences, revealing an author's (or translator's) particular stance in relation to the text. However, "lexical patterns' are a symptom, not a cause, of coherence" (Green and Morgan in Hatim and Mason 1990:194) and our initial choice to adopt one mode of analysis – as opposed to another – heavily conditions our analysis. For example, how we decide to lemmatize particular lexis influences which patterns 'emerge' in our analysis. If we accept that 'quality' reflects the particular nature of a text, as revealed through the

evidence of ‘emerging’ patterns and structures, we must also accept our role in the *construction* of these patterns and – ultimately – in the definition of a text’s structure. More importantly, our interest in lexical items is a feature of their value as signs. The semiotic value of ‘politiche di *welfare*’ [welfare policies] in an Italian context (sentence 9, text sample 6) becomes lost in an ‘automatic’ lexical analysis of this kind, in terms of quality assessment.

Much of our perception of both written and oral texts also relies on those lexical items – encapsulators – which not only repeat concepts, but also communicate the interpersonal (functional-semantic) component of a text. Essential in aiding translators (and interpreters) to ‘read between the lines’ of a text, encapsulation often occurs at the beginning of a paragraph and functions as an organizing principle in discourse structure (Conte 1999: 111). Our corpus contained several examples of this particular use of anaphora, capable of surreptitiously passing an author’s personal assessment of events off as shared information (cf. D’Addio-Colosimo 1988:145).

### 8. Future work

The significance of creating text abridgements depends largely on the functions they serve. We have shown the value of lexical repetition as a cohesive device and illustrated the complex nets lexical items form (after Hoey). We introduced Hoey’s categories of lexical repetition used for automatic processing (tab. 1) and analyzed text 1 from our corpus to identify lexical repetition links with the aid of a concordancer program. We then examined what the process was able to tell us about how cohesion contributes to text coherence (§5.3, §5.4). Our findings were then extended to the assessment of parallel texts in translation studies (§7). In our discussion (§8) we sorted out loopholes found after using Hoey’s model, but also stressed the value of using a concordancer to detect textual cohesion and point to discourse structure.

Ours is a small-scale, exploratory study extended to the assessment of translated texts. At times it would have been easier to bypass the concordancer for a comprehensive analysis, seeing that many of our corpus texts are one-page introductions. However, concordancers prove to be invaluable when examining longer texts and larger corpora, where parallel texts may be aligned for analysis (cf. Zanettin 2000). Yet, a number of ‘missing links’ in our study indicate the need for an extended perspective of analysis, one capable of revealing dynamic discursal activity.

Any theory underlying textual analysis (which may lead to actual ‘rules’ to follow) is based on our experience as text receivers and intercultural experts. Kristeva’s notion of a ‘horizontal axis’ and a ‘vertical axis’ (1980) proposes different perspectives from which to analyze intertextuality and stresses our role as constructive text receivers. Our findings suggest the formulation of a lexical cohesion model – perhaps an extended version, such as the one proposed by Klaudy and Károly – capable of identifying intertextual chains and discourse structures. The program Sphinx Lexica® is able to analyze large corpora in this respect.<sup>10</sup> However, once the ‘rules’ are established and a text is put through a computer program for analysis, what survives may, for example, be a readable abridged version of a longer text, as many of our corpus texts remarkably are. The point is that how an analyst deals with the many shades of grey undoubtedly colouring lexical items and their relationships, when preparing a text for computer processing, heavily conditions what a computer will

do to produce a text abridgement, or how we assess parallel texts. In the final analysis, we as text receivers and producers (translators and interpreters) create our own *nets* to support our own *tightropes*. This implies the need for these nets to stand up to our experience and successfully mirror the cognitive processes underlying our work as translators and interpreters.

#### NOTES

1. In §8 we discuss research carried out by Klaudy and Károly (2000) who also applied an adapted version of Hoey's model of lexical repetition to the assessment of parallel texts.
2. We wish to adapt Hoey's model for use in the analysis of oral texts in simultaneous interpreting, where Gile assumes "interpreters work near saturation level (the 'tightrope hypothesis')" (1999:157, my emphasis).
3. cf. Hatim and Mason, 1990:5-6, 111.
4. cf. Bertalanffy (1950).
5. Appendix 1 lists the volumes in our corpus.
6. University of Bologna library, in Forlì.
7. cf. Sperber & Wilson (1986), *Relevance: communication and cognition*. Oxford: Blackwell.
8. *Concordance*, version 2.0.0, 18 December 2000, copyright © R.J.C. Watt 1999, 2000.
9. cf. Hoey 1997, on the identification of paragraph boundaries, and Hoey 2000, an analysis of hidden lexical clues of textual organization.
10. <<http://www.lesphinx-developpement.fr/en>>.

#### REFERENCES

- BEAUGRANDE, R. DE and W. DRESSLER (1981): *Introduction to Text Linguistics*, London/New York, Longman.
- BERTALANFFY, L. VON (1968): *General Systems Theory*, New York, Braziller.
- CAPRA, F. (1997): *The Web of Life*, New York, Anchor.
- CONTE, M.-E. (1999): *Condizioni di coerenza [Conditions of coherence]*, (new edition, with the addition of two essays edited by B. Mortara Garavelli), Alessandria, Edizioni dell'Orso.
- D'ADDIO-COLOSIMO, W. (1988): "Nominali anaforici incapsulatori: un aspetto della coesione lessicale" [*Nominal anaphoric encapsulators: An aspect of lexical cohesion*], in T. DE MAURO *et al.* (eds.), pp. 143-151.
- GILE, D. (1999): "Testing the Efforts Models' tightrope hypothesis in simultaneous interpreting – A contribution," *Hermes, Journal of Linguistics* 23, pp. 153-172.
- GREEN, G., and J. MORGAN (1981): "Pragmatics, grammar and discourse," in P. COLE (ed.), *Radical Pragmatics*, New York, Academic Press.
- HALLIDAY, M.A.K. and R. HASAN (1976): *Cohesion in English*, London/New York, Longman.
- HATIM, B. and I. MASON (1990): *Discourse and the Translator*, London/New York, Longman.
- HOEY, M. (1991): *Patterns of Lexis in Text*, Oxford, OUP.
- HOEY, M. (1997): "The interaction of textual and lexical factors in the identification of paragraph boundaries," in MECHTHILD, R. and W. THIELE (eds.), pp. 141-167.
- HOEY, M. (2000): "The hidden lexical clues of textual organisation: A preliminary investigation into an unusual text from a corpus perspective," in BURNARD, L. and T. MCENERY (eds.), pp. 31-39.
- KÁROLY, K. (1999): "An analytical tool for the study of lexical repetition," *Modern Filológiai Közlemények* 1-1, pp. 40-59.
- KLAUDY, K. and K. KÁROLY (2000), "The Text-organizing Function of Lexical Repetition in Translation," in OLOHAN, M. (ed.), pp. 143-159.
- KRISTEVA, J. (1980): *Desire in Language: A Semiotic Approach to Literature and Art*, New York, Columbia University Press.
- MASON, I. (2001): "Translator behaviour and language usage: Some constraints on contrastive studies," *Journal of Linguistics* 26, pp. 65-80.

- MATURANA, H. R. and F. J. VARELA (1980): *Autopoiesis and Cognition*, Dordrecht, Reidel.
- MATURANA, H. R. and F. J. VARELA (1998): *The Tree of Knowledge: The Biological Roots of Human Knowledge*, Boston/London, Shambhala.
- OLOHAN, M. (2000): *Intercultural Faultlines: Research Models in Translation Studies*, Manchester, St. Jerome Publishing.
- SEIDLHOFER, B. (2000): "Operationalizing intertextuality: Using learner corpora for learning," BURNARD, L. and T. McENERY (eds.), pp. 207-223.
- SPERBER, D. and D. WILSON (1986): *Relevance: Communication and Cognition*, Oxford, Blackwell.
- WIDDOWSON, H.G. (1978): *Teaching Language as Communication*, Oxford, OUP.
- ZANETTIN, F. (2000): "Parallel corpora in Translation Studies: Issues in corpus design and analysis," in OLOHAN, M. (ed.), pp.105-118.

## APPENDIX 1

### Volumes containing corpus texts

1. Deacon, Bob, Hulse, Michelle & Stubbs, Paul 1997. *Global Social Policy*, London: Sage.
2. Henig, Stanley 1997. *The Uniting of Europe*, London/New York: Routledge.
3. Holloway, David & Sharp, Jane M. O. (eds) 1984. *The Warsaw Pact: alliance in transition?*, London: Macmillan.
4. Mayne, John & Zapico-Goñi (eds.) 1997. *Monitoring Performance in the Public Sector*, New Brunswick, N.J.: Transaction.
5. McInnes, Colin (ed.) 1992. *Security and Strategy in the New Europe*, London/New York: Routledge.
6. Mearsheimer, John J. 1983. *Conventional Deterrence*, Ithaca/London Cornell University Press.
7. Raz Joseph 1986. *The Morality of Freedom*, Oxford: OUP.
8. Sagan, Scott D. & Waltz, Kenneth N. 1995. *The Spread of Nuclear Weapons: a debate*, New York/London: Norton.
9. Sarangi, Skrikan & Slembrouk, Stefaan 1996. *Language, Bureaucracy and Social Control*, Foreward by Norman Fairclough, London/New York: Longman.
10. Taylor-Gooby, Peter (ed.) 1998. *Choice and Public Policy: the limits to the welfare markets*, London: Macmillan.