

Construction of Thematic Representations of Texts Based on Domain-Specific Thesaurus

Natalia V. Loukachevitch and Boris V. Dobrov

Research Computing Center of Moscow State University; NCO Center for Information Research
339, Research Computing Center of Moscow State University, Vorobyevy Gory, Moscow, 119899, Russia
{louk, dobroff}@mail.cir.ru

Abstract

The paper considers interrelations between lexical cohesion and the thematic structure of a text. The technique of automatic construction of the thematic representation of the text contexts is described. The technique uses knowledge from Sociopolitical thesaurus, which was specially developed as a tool for automatic text processing.

Introduction

The information access domain includes various applications such as retrieval of documents, text categorization, information filtering, text summarization and others. Automatic systems in any such subdomains can be much more effective if their text processing is based on hierarchical structure of the themes and subthemes, or at least if the thematic significance of terms in texts is known.

One can see that when a term is very important for a text, when it belongs to the main theme of a text, then the whole text contains a lot of words semantically related to the term. Therefore it is natural to suppose that if one can find such sets of related terms it can help to identify the main theme and the subthemes of documents.

The phenomenon of semantically related words in a text is usually considered as lexical cohesion phenomenon (Halliday and Hasan 1976). Lexical cohesion is the most frequent type of cohesion in a connected text. It can be expressed by repetitions, synonyms and hyponyms or by words connected with other semantic relations such as whole - part, situation - participant, object - property and so on. Halliday and Hasan supposed that pairs of words connected by cohesive relations form lexical chains

Several authors proposed their approaches to constructions of lexical chains based on such linguistic resources as WordNet (Miller et al. 1990) and Thesaurus Roget's and used them in such applications as information retrieval, text summarization, correction of spelling errors and others.

But the implemented systems met real problems.

In (Hirst and St-Onge 1998) the most important problems in automatic lexical chain construction were formulated as follows:

- (1) limitations in the set of relations in WordNet, or a missing connection; especially lack of situational relations (school - child care, physician - hospital);
- (2) lack of relations between different parts of speech;
- (3) inconsistency in the proximity implicit in links in WordNet (*stew* and *steak* were not considered as related because distance of 6 synsets; *public* and *professionals* are considered as related - distance 4 synsets).

In this paper we will argue that the problems of the lexical chain approaches are only partially due to quality and structure of the used linguistic resources. The proposed approaches do not account important properties of the text structure.

To overcome the problems we introduce new notions "thematic node" and "thematic lexical chain". A thematic node is a conceptual structure, providing connections between different levels of the thematic structure of a text, a thematic lexical chain is a textual representation of a thematic node. We will propose the technique of constructing the thematic representation of a text as a hierarchical structure of thematic lexical nodes.

Our technique is based on bilingual Thesaurus on Sociopolitical life (Loukachevitch, Salii, and Dobrov 1999), which contains 27 thousand concepts, 62 thousand Russian terms, 47 thousand English terms. We consider the sociopolitical domain as a domain of events and problems which are significant for society as whole. These problems are discussed in official documents and newspaper articles, in legislative acts and international treaties. So the thesaurus comprises terms from economic, political, social, legislative, cultural, sports and other spheres.

The Thesaurus is a main tool of our research, and we have to indicate its main features. The thesaurus was created specially as a tool for automatic processing of large text collections. Therefore its concepts have detailed rows of synonyms and variants. Descriptions of concepts include much encyclopedic information: possible situations, participants, properties and so on, which may be useful for the information retrieval tasks. Types of conceptual relations

originate in the relations of conventional information-retrieval thesauri (broader term – narrower term, related term). In our thesaurus broader-narrower relations can be marked with modifiers to describe multiple inheritance. Most of the symmetric Related-Term relations of conventional information retrieval thesauri are described as non-symmetric dependency relations. Dependency relations describe how existence of a concept (or examples of a concept) depends on existence of another concept (or its examples) (Guarino 1998). The thesaurus relations have such properties as symmetry, transitiveness, inheritance.

Automatic Construction of Lexical Chains

Recent works simulated lexical cohesion, constructing 'lexical chains'. A lexical chain is a chain of words in which the criterion for inclusion of a word is some kind of cohesive relationship to a word that is already in the chain (Morris and Hirst 1991). Morris and Hirst also proposed a specification of cohesive relations based on Roget's Thesaurus.

Hirst and St-Onge (1997), Barzilay and Elhadad (1997) construct lexical chains based on WordNet relations. Definition of lexical cohesion relations based on WordNet includes extra-strong, strong and medium-strong relations. Every next relation is weaker than previous one. Medium-strong relations include paths of the WordNet conceptual structure with maximum 5 links and have a specific form: "UP--...--UP", "DOWN--...--DOWN", "UP--... --...--DOWN", "UP--...--HORIZONTAL--...--DOWN".

The main stages in the proposed construction of lexical chains are as follows:

- the construction of lexical chains begins from the first words of a text;
- to insert the next word, its relations with members of existing lexical chains are checked.

If there are such relations with any element of a chain then the new word is inserted in the chain. Only one lexical chain can be chosen. Among several possible lexical chains a lexical chain with maximal weight of a relation with a current word is chosen. For strong relations and medium-strong relations there are restrictions on distance between a current word and existing lexical chains.

We tried to repeat the described process with human experts. An expert received sentence by sentence from a text and had to construct lexical chains. The main contents of a text was not known to an expert. In this process any decision on nontrivial connection (not repetitions, hyponyms and hypernyms) was a real problem. When the main contents of a text was known to experts (after several times of rereading), then it seemed to be much easier to make decisions.

A human needed overall information about the text contents, but the algorithm do not use any information about the whole text. In fact, it looked like a circle. We need lexical chains to find the main theme of a text, but the

main theme considerably influences process of lexical chain construction.

There is another problem. What happens if this algorithm has additional situational relations. Let us take text 3 from text collection of Barzilay used in experiments on lexical chain construction (www.cs.columbia.edu/nlp/summarization-test/index.html). The beginning of this text is as follows:

Example1. *Partying post-op (Text3)*

COMPANIES that announce breakthroughs in medical science usually get a hero's welcome. But whereas surgeons applauded when Health Care International (HCI), a private hospital, announced this month that one of its doctors had invented an anaesthetic machine which means people feel much less sick after operations and speeds their recovery, reaction in Scotland was less than ecstatic.

In 1994, HCI opened a swanky 240-bed private hospital to do expensive surgery on rich foreigners at Clydebank in Glasgow. Though it promised hundreds of jobs in an area of long dole queues, it went straight to the top of the local hate list...

We take deficient situational relations from Thesaurus on Sociopolitical life (Loukachevitch, Sali, and Dobrov 1999). Situational relations described in the Thesaurus can be hierarchical or horizontal. Therefore we can repeat exactly the same algorithm of constructions of medium-strong relations as described.

In the first paragraph of the text we can see the following terms from medical domain: *medical science, surgeons, health care, hospital, doctor, anaesthetic, sick, operation*. In the Thesaurus we can find the following paths of relations, consisting of hyponymy-hypernymy relations and situational relations. All forms of the paths are allowable by the described algorithm:

Surgeon -- UP -- *doctor* -- UP -- *medical field* -- DOWN -- *medical science*;

Health care -- DOWN -- *medical field* -- DOWN -- *medical science*;

Hospital -- UP -- *medical organization* -- UP -- *medical field* -- DOWN -- *medical science*;

Doctor -- DOWN -- *surgeon*;

Anaesthetic -- UP -- *medical help* -- UP -- *medical field* -- DOWN -- *medical science*

Sick -- UP -- *health problems* -- DOWN -- *medical field* -- DOWN -- *medical science*

Operation -- UP -- *surgery* -- DOWN -- *surgeon*.

Existence of such paths means that all medical terms of this text are be elements of the same lexical chain. All terms are situated very close to each other. Differences in weights of cohesion relations do not work, because there is no alternatives. But such a lexical chain does not correspond to the main theme of the text.

An expert proposed the following formulation of the main theme: "A private hospital Health Care International invented an anaesthetic machine, which allows patients

feel much less sick after operations". So the situation, described in the text, includes at least four important "medical" participants, playing quite different roles in the situation: *hospital, anaesthetic, patient, operation*. Therefore the results of the lexical chain construction misrepresent the thematic structure of the text.

In fact, the same problem can appear in processing of any text within any domain. Let us see, for example, next text of the collection "A shortage of sirs" discussing "shortage of teachers at schools of Britain".

The beginning of the text is as follows:

Example 2. Teaching

A shortage of sirs

TRYING to teach quadratic equations to a bunch of unruly teenagers is hardly the cushiest of jobs. So as unemployment continues to fall, it is not surprising that teacher-training colleges are finding it hard to recruit prospective maths teachers. But there are growing worries that the supply of teachers may be dwindling just as the number of school-age children is growing...

Again *teacher* and *school* are very related words. But in this text they have to be represented as different entities in the thematic structure, because the relation between them is in the focus of the text.

Thus the approach does not explain two interdependent phenomena:

1) why experts need information about the whole text to identify cohesion relations;

2) why very related terms are separated in the thematic structure of some texts.

Thematic Structure and Thematic Lexical Chains

The described approach to lexical chain construction searches relations between sentences of a text. But such relations can be as far from the thematic structure of a text as superficial syntactic relations from deep semantic relations. However, the primary goal of the most information retrieval applications is to recover the thematic structure of texts and to use it for effective text processing. Therefore it is necessary to analyze facts known about the thematic structure.

Van Dijk and Kintsch (1983) describe the topical structure of a text, the macrostructure, as a hierarchical structure in a sense that the theme of a whole text can be identified and summed up to a single macroproposition. The theme of the text can be usually described in terms of less general themes which in turn can be characterized in terms of even more specific themes, and so on. Every sentence of a text corresponds to a subtheme of the text.

This means that a connected text has its main theme and this main theme can be formulated. Let us take the Example1 and the formulation of its main theme (see previous section). The formulation of the main theme names the

most important concepts or objects of the text (the main concepts of the text) such as *hospital, machine, anaesthetic, patient, operation* and indicates relations between them. (We imply that the thematic structure is a structure of concepts and specific objects, not words. Therefore here and below when we say 'concepts of a text' we imply concepts, the text variants of which were mentioned in a text).

The subthemes of the texts have to describe some of the main concepts or relations between them. Therefore if it is required to find the subthemes of a text it is necessary to formulate the main theme of a text, identify the main concepts of the text; construct singles, pairs, triples from the main concepts as names of the subthemes and find corresponding sentences.

So all main concepts from the formulation of the main theme of the Example1 are supported with several sentences.

For example, there is a subtheme and corresponding sentences describing the hospital: *In 1994, HCI opened a swanky 240-bed private hospital to do expensive surgery on rich foreigners at Clydebank in Glasgow...*

Another subtheme describe relations *anaesthetic -- machine -- patient*:

1)... *one of its doctors had invented an anaesthetic machine which means people feel much less sick after operations and speeds their recovery.*

2)*But Mr Kenny's device uses two micro-computers to achieve much closer monitoring of the pump feeding the anaesthetic into the patient and others.*

At the same time we can find several sentences and clauses, presence of which in the text is difficult to explain on the basis of the proposed formulation of the main theme:

1) *reaction in Scotland was less than ecstatic.*

2) *Though it promised hundreds of jobs in an area of long dole queues, it went straight to the top of the local hate list.*

3) *The Edinburgh house in which Sir James Young Simpson, another Scottish doctor, discovered the first modern anaesthetic, chloroform, in 1846 has a commemorative plaque on it. Mr Kenny may eventually be as esteemed as Sir James. But whether his innovation encourages the Scots to pay more regard to HCI remains to be seen.*

These clauses and sentences describe relation of Scottish people to the hospital and the invention. It allows us to suppose that such important concept as SCOTLAND was missed in the formulation of the main theme.

Thus every sentence has correspondence with a subtheme, every subtheme has a corresponding fragment in a formulation of a subtheme of upper level or the main theme.

To refer to the main theme a subtheme has to include a main concept or its related concept; in sentences of a text such references look like lexical cohesion relations. For example, the subtheme *anaesthetic -- machine* includes the

following sentence with reference words *device* and *micro-computer*:

But Mr Kenny's device uses two micro-computers to achieve much closer monitoring of the pump feeding the anaesthetic into the patient.

Word *anaesthetists* refers to *anaesthetic*: *Anaesthetists who have either tried the machine, or seen it demonstrated at medical conferences, are enthusiastic.*

Therefore sets of lexically related words in a text play a very important role. They provide references from the lower levels of the thematic structure of a text to the upper levels of subthemes and to the main theme. Every such reference has its specific referent – a concept of upper level.

On the other hand every main concept or a concept of upper levels of the thematic structure has a set of related concepts, text variants of which were used for reference on it in a text. Such a set looks like a node. It has its center and all other its concepts are semantically related to the center. Therefore we called such sets “thematic nodes”. Thematic nodes have inseparable relations with the thematic structure of a text and definite functionality within it. So the concepts of a text can be divided to thematic nodes.

In a text the text variants of the same thematic node look like sets of semantically related words but the main property of words in such sets is their relatedness to the same concept. Textual representation of thematic nodes is called “thematic lexical chain”. A thematic node and a thematic lexical chain that have a main concept of a text as a thematic center are called “main thematic node” and “main thematic lexical chain”, respectively.

Thus, in our opinion, sentences in a coherent text can be superficially connected by various means such as pronouns, cue words, pairs of related words. But sets of related words in different sentences of a text, which look like lexical chains, have tense functional relations with the thematic structure – they provide connections between different levels of the thematic structure. Such sets have their own internal conceptual form – form of thematic nodes.

Properties of Main Thematic Nodes

From the consideration of relationships “the main theme -- subthemes -- sentences -- related words” two important consequences follow. At first we can formulate a very important principle how to find main concepts of a text, that is how to distinguish main thematic nodes from other thematic nodes.

The main theme of a text usually includes several main concepts. The text discusses relations between the main concepts, that is, there are subthemes which devoted to discussing a specific relation. Therefore combinations of such main concepts and their related concepts have to often occur together in clauses and sentences of the text. They have to occur together more often than concepts belonging to other thematic nodes. If there are n main concepts in a

text, then the sum of frequencies of mutual cooccurrence of concepts of different main thematic nodes have to be more than sum of frequencies of other n thematic nodes. Otherwise the formulation of the main theme was not correct.

This feature of main thematic nodes allows their recognizing among thematic nodes of other terms for texts of any size and different genres.

As the second consequence we can formulate a very important restriction on the membership in a thematic node. Two concepts of a text (even very semantically related concepts) can not be considered as elements of the same thematic node if they often co-occur in the text. Indeed, frequent cooccurrence means that a relation between the concepts is an important point of discussion in the text. If frequency of their mutual cooccurrence is maximal or comparable with maximum cooccurrences of other pairs of concepts in the text, it possibly means that the relation between the concepts is a focus of the whole text, and both concepts C1 and C2 are main concepts of the text. The relatedness of concepts does not disappear. But the concepts are considered as different entities contrapositive to each other. Therefore it is impossible to gather these concepts in the same thematic node, otherwise the thematic structure of the text is distorted.

Such situation can be seen in Example2. Concept *teacher* was mentioned 24 times. Concept *school* – 9 times. At least 7 times they were mentioned near each-other. This is maximum frequency of mutual cooccurrence for the text, and the both concepts were in the main theme of the text.

In the Example1 pair *hospital* – *patient* (which often considered as a pair of related concepts) coocurs 3 times, pair *anaesthetic* – *patient* 4 times (the maximum frequency is 6 of pair *patient* – *machine*). So these values of cooccurrence are significant for the text

So in this examples we can really see interdependence between frequency of cooccurrence of two concepts in a text and the thematic structure of a text.

Thus our conclusion is as follows: the more frequency of mutual cooccurrence of two concepts in a text, the more importance of discussion of the relation between the concepts, the more necessity to include the concepts in different thematic nodes. And vice versa if frequently mentioned concepts are relatively rarely met near each other, it can evidence in favor of their semantic relatedness and necessity to include them in the same thematic node.

The main thematic lexical chains in the Example1 are:

Machine (8): *pump, device, computer, micro-computer, equipment, medical equipment*

Anaesthetic (7): *anaesthetist, anaesthesia, drug, chloroform*

Patient (10): *sickness, sick, people*

Scotland (1): *Glasgow, Scottish, Scots, Britain, British*

Hospital (6): *HCI*

Operation (3): *surgery, surgeon, post-operative.*

This set of main thematic lexical chains is a result of manual analysis. In automatic processing relations *machine*

– *medical equipment, anaesthetic – chloroform, operation*
– *surgeon, hospital – HCI* were missed.

Practical Steps to Creation of Thematic Nodes. Thematic Representation

The main stages of automatic constructions of thematic nodes for a text, implemented in a real system are as follows:

1) To construct thematic nodes it is not necessary to follow pairs of related words sentence by sentence. At first it is necessary to identify suppositional centers of the thematic nodes. The main concept of a thematic node are relatively more important than other concepts of the thematic node and has to be stressed in a text. It can be more frequent than related concepts, can be used in the title or subtitle of the text or be mentioned before other related concepts.

A main concept is a member of the only thematic node. Other concepts can be included in several thematic nodes (in practical realization we used restriction “three thematic nodes” as maximum). So concept *anaesthetist* in the Example1 refers not only to concept *anaesthetic* but also related to concept *doctor*.

All concepts of a text are ordered by their initial weight (frequency, title). Every next concept that is not a member of an existing thematic node can become a thematic center of a new thematic node.

2) The chosen main concept collects all semantically related concepts mentioned in the text. Definition of semantic relatedness is based on allowable thesaurus paths with several restrictions. The restriction described in section3 is not yet implemented but it is simulated with restrictions in thesaurus paths (paths up-...- down are not permitted). At the same time it is important to stress that presence of situational relations in the Thesaurus provides much more straight paths than in WordNet 1.6.

3) To estimate cooccurrence of concepts of different thematic nodes in a text, we use the notion 'textual relation': a given concept has textual relations with those concepts of the text that are located no further than three concepts of the Thesaurus from the given concept (location order is not important).

Thus the context of a concept occurrence is determined by the quantity of meaningful elements. Other words (not from the Thesaurus) are not included in the count. Textual relations pass through sentence borders and are interrupted only by paragraphs.

4) In our approach the main thematic nodes are those ones that:

- have textual relations with all other main thematic nodes and
- have a sum of frequencies of textual relations between these nodes that is greater than the sum of frequencies for the same number of other thematic nodes of this text.

More detailed description of the algorithm can be read in (Loukachevitch and Dobrov 2000).

We incorporate constructed thematic nodes into the 'thematic representation' of the text. The thematic representation of a text is a hierarchical structure of concepts where concepts semantically related to thematic centers are gathered in thematic nodes. Thematic nodes whose thematic centers can characterize the contents of a text are main thematic nodes. A hierarchy of thematic representation characterizes the importance of concepts in the text: the thematic center is more important than other concepts of the thematic node; concepts of main thematic nodes are more important than the concepts of other thematic nodes.

Thus all concepts of the text are divided into five classes of different importance for the text:

- main concepts of main thematic nodes,
- other concepts of main thematic nodes,
- main concepts of specific thematic nodes,
- other concepts of specific thematic nodes,
- mentioned concepts.

More detailed description of the algorithm can be read in (Loukachevitch and Dobrov 2000).

Evaluation of Applications Based on Thematic Representation

We tested our approach for concepts of a text corresponding to the main theme of the text. Thus, we could test how information, described in the Thesaurus, supported exposition of the main themes of various texts.

For every text we tried to choose three or four main concepts characterizing the main theme of the text in the best way. We chose them mainly from the title, the first paragraph of the text or took the most frequent concepts manually.

At the second stage for each main concept we fulfilled automatic procedure of construction of a thematic node.

Finally during manual reading we tested if every element of the thematic node really served for establishing cohesive relations with the its thematic center. In this process we could compute precision and recall of the automatic process of detection of lexical cohesion relations in texts.

After analysis of 73 main thematic nodes, serving for organizing lexical cohesion relations in 25 texts of the sociopolitical domain, our results are as follows: precision – 89 %, recall - 71%.

The Thesaurus and the thematic representations of texts are basic tools for organizing conceptual indexing and ranged information retrieval, automatic text categorization and text summarization in University information system RUSSIA (www.cir.ru/eng), containing 300 thousand Russian documents. Calculation of weights of concepts in texts is based on automatically constructed thematic representations of the texts.

To evaluate thesaurus-based information retrieval in UIS RUSSIA we took 20 topics from list of “Subject Headings for Legislative Acts” adopted as an official system of sub-

ject headings in the Russian Federation, includes 1168 subject headings and 20 main thematic subdivisions. The topics for evaluation were randomly extracted from every thematic subdivision (for example, “Water supply”, “Use of nuclear energy”, “Migration of population”).

Every search was implemented twice. The first search was implemented using vector model. In the second search we manually represented a topic as Boolean expression of words which are absent in the Sociopolitical thesaurus and terms from the Thesaurus. Such translation was literal without any additions or deletions. For example, subject heading “Use of nuclear energy” was represented as “*use (word)*” and “*nuclear energy (term)*”.

During search every term was automatically expanded using its full thesaurus tree of hierarchy.

We used “3-point” evaluation to compute average precision for 0.2, 0.5, 0.8 recall values (Vorhees 1998), as a single measure of retrieval effectiveness. Results of evaluation were as follows:

Type of search	0.2	0.5	0.8	Avg
Vector Search	0.77	0.52	0.02	0.44
Thesaurus Search	0.81	0.58	0.46	0.62

Coverage of media texts by the thesaurus terms allowed our text summarization system to receive qualitative results in SUMMAC conference in 1998 (SUMMAC Final report 1998).

Acknowledgements. Partial support for this research is provided by the Russian Foundation for the Humanities through grant # 00-04-00272.

Conclusion. Challenge questions

The thematic representation model does not depend on genre, size, domain or language of texts. It is “only” necessary to change a thesaurus if a domain or a language has been changed.

The thematic representation is very illustrative. We processed thousands of texts, looked through their thematic representations, in order to test how the thematic representations correspond to the main themes of texts. Misrepresentation of the main theme in the thematic representation usually was due to mistakes in the thesaurus descriptions. At the first stages it was the most important factor for correcting the thesaurus and even the thesaurus model.

For several years the thematic representation model and the Thesaurus are intensively used in various applications of automatic text processing of large text collections in Russian and English, such as automatic conceptual indexing, information retrieval, knowledge-based text categorization and automatic summary construction.

In future we plan to study how to incorporate rhetorical relations (Mann and Thompson 1988), resolution of pronoun anaphora, partial syntactic analysis into our technique of analysis of the thematic structure of coherent texts.

References

Barzilay, R. and Elhadad, M. 1997. Using Lexical Chains for Text Summarization. In Proceedings of ACL/EACL Workshop Intelligent Scalable Text Summarization. Madrid.

Guarino, N. 1998. Some Ontological Principles for Designing Upper Level Lexical Resources. In Proceedings of First International Conference on Language Resources and Evaluation.

Halliday, M. and Hasan, R. 1976. *Cohesion in English*. Longman, London.

Hirst, G. and St-Onge, D. 1998. Lexical Chains as representation of context for the detection and correction malapropisms. In Fellbaum C., editor, *WordNet: An electronic lexical database and some of its applications*, 305-332. Cambridge, MA: The MIT Press.

Loukachevitch N. 1998. Text Summarization Based on Thematic Representation of Texts. In: Intelligent Text Summarization - Proceedings of the AAAI'98 Spring Symposium on Intelligent Text Summarization - AAAI Technical Report SS-98-06, pp.77-84.

Loukachevitch, N.; Sali, A. and Dobrov, B. 1999. Thesaurus for Automatic Indexing: Structure, Development, Use, In *Proceedings of International Congress "Terminology and Knowledge Engineering"*, 343-355.

Loukachevitch, N., and Dobrov, B. 2000. Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems. *Machine Translation Review* 11: 10-20. (see also www.bcs.org.uk/siggroup/nalatran/mtrreview/mtr-11/mtr-11-8.htm)

Mann, W. and Thompson, S. 1988 Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8(3): 243-281.

Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; and Miller, K. 1990. Five papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton University.

Morris, J. and Hirst, G. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of a text. *Computational Linguistics* 17(1): 21-48.

The TIPSTER SUMMAC Text Summarization Evaluation. Final Report 1998. MITRE Technical Report MTR 98W0000138

Vorhees, E. 1998. Using WordNet for Text Retrieval. In Fellbaum C., editor, *WordNet: An electronic lexical database and some of its applications*, 285-303. Cambridge, MA: The MIT Press.