Pergamon

Language Sciences, Vol. 19, No. 4, pp. 391–424, 1997

0388–0001/97 $17.00+0.00

PII: S0388–0001(97)00003–X

WHAT IS TECHNICAL TEXT?

TERRY COPECK,[a] KEN BARKER,[a] SYLVAIN DELISLE,[b]
STAN SZPAKOWICZ[a] and JEAN-FRANÇOIS DELANNOY[c]

Beyond labeling it easier to process than other types, few researchers who use technical text in their work try to define what it is. This paper describes a study that investigates the character of texts typically considered technical. We identify 42 features of a text considered likely to correlate with its degree of technicality. These include both objectively verifiable measures like *marked presence of interrogative or imperative sentences* which are akin to the criteria used by Biber in *Variation Across Speech and Writing*, and subjective measures such as *presence of hierarchical organization*. All are less ambiguous than technicality, so our inventory may be suited to use in a procedure that classifies text as technical or non-technical. An inventory organizing and describing these lexical, syntactic, semantic and discourse features was used to rate nine varied sample texts. Analysis of 22 ratings of each text indicated that 31 features in the inventory were meaningful predictors of text technicality when considered independently. The inventory has been revised and a formula to compute technicality has been developed in the light of these findings. © 1997 Elsevier Science Ltd. All rights reserved

1. Introduction

Of all the descriptions that are applied to writing, one of the most notorious is the term *technical*.[1] When someone calls language technical in everyday life, it is usually understood to mean that its message is hard to grasp. In scientific circles however the modifier has a happier connotation. The expressiveness and flexibility of natural language makes it hard to process in a methodical fashion, and researchers generally understand 'technical text' to mean writing which is more tractable because, for example, it lacks figurative language and can be understood in its literal sense.

Notwithstanding its usefulness to researchers, technical text does not appear to be a well-defined term. A recent search on the Internet found it used in the following ways:

- besides ourselves, *Natural language processing researchers* at Mississippi State University have been working for a number of years on the problem of extracting information from technical text. The texts in question were the Wampole Veterinary Handbook and more recently, Chemical Abstracts;
- *social scientists* have constructed the Aston Scientific and Technical English Corpus (ASTEC), which describes itself as 'a facility at the disposal of staff and research students wishing to investigate the nature and characteristics of natural language', and lists 'discourse features in scientific and technical text' among the research fields it is suited to;
- *educators* talk of teaching students in technical schools and colleges to produce clear technical writing;

[a]Department of Computer Science, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5 {kbarker, terry, szpak}@csi.uottawa.ca 613-564-5420 fax: 613-564-9486
[b]Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada G9A 5H7, Sylvain_Delisle@uqtr.uquebec.ca 819-376-5125 fax: 819-376-5185
[c]RES International Inc. 600–100 Sparks Street, Ottawa, Ontario, Canada K1P 5B7, delannoy@res.ca, 613-238-3690 fax: 613-235-589

- a *survey author* uses the term as a classification category in a survey on usage of the web;
- the *commercial* Logos Corporation describes their server as capable of quickly translating English or German technical text into a number of other languages. Firms like typesetters and documentation producers advertise their subject matter as being technical text.

Other users go further and provide some indication of what they consider technical text to be:

- two *researchers into hypertext*, describing how readers quickly assess a book by flipping through its pages, speak of 'the technical content (e.g. number of equations, number of technical drawings)';
- a *contributor to a Usenet newsgroup* on OCR says, 'By technical text I mean text with complicated mathematics, typically as found in science textbooks' in a message about computerizing text recognition;
- the *authors* of *Machine Translation: An Introductory Guide* state that technical texts 'typically contain large numbers of terms'.

It is clear that many people employ this notion and it is likely that their understandings of it do not accord with one another. This situation, which is essentially unchanged from years past, motivated the project discussed in this paper. Our concern is knowledge extraction from natural language and other forms of language processing and, like others listed above, we limit our efforts to technical text. The work discussed here is an attempt to investigate just what that term involves.

2. Identifying the task

We began work with the general intention of investigating the nature of technical text and the particular objective of finding practical ways of determining whether a given text is technical or not. Our appreciation of what is involved developed over the course of the study into relatively firm views on a few issues. These might be termed the assumptions of our study.

2.1. Assumption

Text technicality rather than technical text. Although the application of any categorical mechanism would be to classify a given text as either technical or not, yes or no, technicality itself appears to be a *graded* rather than a binary quality, with individual texts possessing more or less of it than one another and with its exact measure in some texts hard to determine. Individual researchers will set the bar of acceptance and rejection at whatever point along the underlying technicality grade best suits their purposes; and whatever it is set, some texts will always be borderline and hard to call. One corollary of this fact is that it is easier to rate texts on their technicality in *relative* rather than absolute terms. That is to say, some people may fairly easily order a set of texts on increasing technicality but find it difficult to give an assessment of any text's technicality in isolation.

Limitations of the process. Unlike a count of the number of characters in a text, technicality is a qualitative feature; and because it is qualitative, it is also subjective. Many (though not all) of the textual features which were found to be correlated with technicality in this study are also qualitative. The study involved readers rating different texts on a number of criteria. In such cases the quality of the rating is only as good as the performance of the rater, and that performance varies between raters and even with any single rater across time. One corollary of

this fact is that any attempt to provide a definitive characterization of technicality is an inherently contentious undertaking. We are not going to try to give one. What we will try to do however is to shift what is assessed in any rating process from the rather abstract notion of technicality *per se* onto one or more overt features of the text, some of which may be countable.

If measures of technicality are subjective, at the same time they can be held to be fairly broadly based. 'Technical text' is not a term defined in a theory. It is a notion in everyday use, sufficiently common that even academic writers do not generally feel they need to provide an explicit definition. Does their exact usage of the term denote the same concept as the everyday usage? In the absence of any evidence to the contrary we assume so. One corollary of this view is that, given a sufficiently large sample, everyday users are a good oracle for measures of technicality; and that is how they have been used in this study.

Scope of application. Although more text is becoming available on-line with each passing day, for the foreseeable future the proportion of all written material which is digitally encoded can only be a small, perhaps minuscule, percentage of all text. The inventory we wish to construct will not be limited to texts which are stored in digital form. It is instead intended to be applicable to text in any human-readable format.

The nature of technicality. What does it mean for a text to be technical? A few paragraphs back it was suggested that technical text is generally understood to contain little or not figurative language. Just above we further characterize technicality as a *graded quality*. What else can be said about technical text that would meet with general acceptance? One way to answer this question is to adopt some larger framework into which technical text would fit as a category. Criteria used to distinguish categories from one another in such an organization would then accrue as appropriate to the category of technical text. Such an exercise is described in the next subsection.

Text genre, text type. Linguists categorize written materials according to *genre*. Some authors (Karlgren, 1995; Karlgren and Cutting, 1994) seem to use this term and *text type* interchangeably. The *Longman Dictionary of Applied Linguistics* defines genre as

> a particular class of speech events which are consisted by the speech community as being of the same type. Examples of genre are: prayers, sermons, conversations, songs, speeches, poems, letters, and novels. They have particular and distinctive characteristics. . .

A typology or taxonomy is an exhaustive classification of individuals into usually disjoint terminal categories. The definition of genre does not appear necessarily to be as rigorous or partitive as this.[2] Is technical text a genre, or a type, or something else? This is a question to which we sought an answer in the literature. We came to the conclusion that technicality is not a genre; that it is instead a feature of a number of genres. And although it can be used as a basis for categorization for this reason—technical versus non-technical—our conviction, after looking into the various ways that others have classified texts, is that technicality is not particularly well suited to serve as a classifier.

As our insight into the character of the phenomenon under investigation deepened, the definition of the task we set ourselves gradually evolved until the objective became to identify a set of the overt features of a text which correlate highly with its technicality. Achieving this would go some distance toward making the assessment of text technicality methodical and, perhaps, able to be automated.

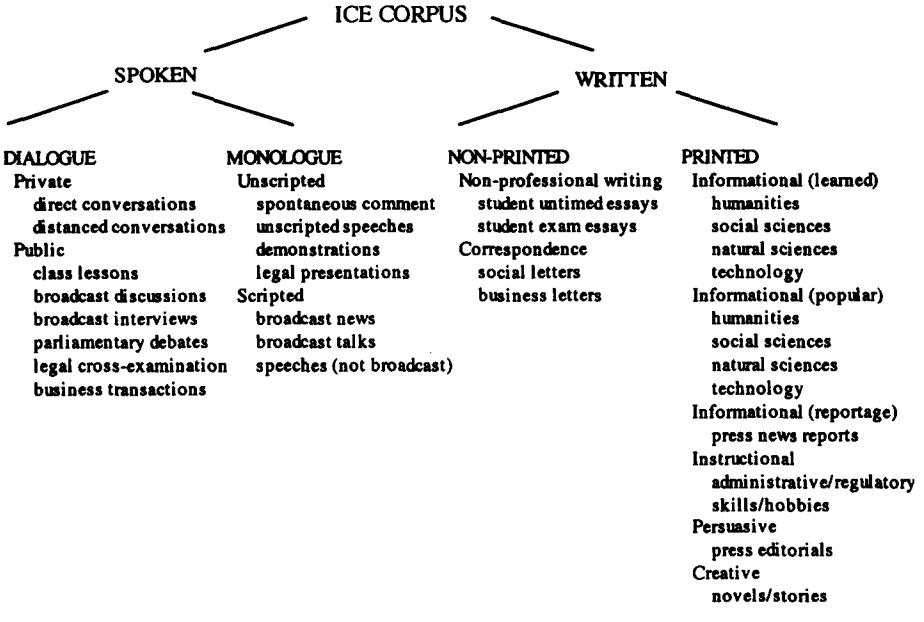

Fig. 1. Categories in the international corpus of English.

2.2. Categorizing written material

Library classification schemes such as the Dewey Decimal and Library of Congress systems are obvious candidates to provide frameworks into which to categorize writing. After some investigation—library science does concern itself with the theory of categorization—these were however set aside because they are primarily topical, and we seek to categorize texts in a more fundamental way. The organizations used in language corpora are more promising; the Brown corpus has been used in research which is relevant to our task. An even more comprehensive organization of all language forms can be found in the International Corpus of English produced by the Survey of English Usage at University College, London. The categories in this corpus appear in Fig. 1.

They give rise to the taxonomy in Table 1. In it 'Printed' and 'Non-Printed' have been replaced by 'Public' and 'Private', respectively,[3] and the literal/figurative distinction, which is so prominent in descriptions of technical texts, has been applied to printed writing. The categorial criteria on the right are taken from the *Longman Dictionary of Applied Linguistics* where available and apply to typical instances rather than to all.

Literal public writing, shaded in the table, would then be roughly synonymous with technical text. Concatenating the characterizations of ancestors of this class in the taxonomy gives minimal general characteristics of this language category:

> A symbolic record on a medium in a generally-understood system of communication employing glyphs and the rules to combine them. These broadly disseminated records are intended for a general readership; they employ words in their literal senses and assume only objective, generally available knowledge about the world and the domain of discourse.

This description provided initial criteria to identify candidates whose technicality is worth evaluating. Texts not satisfying these requirements are unlikely to be technical; however,

Table 1. An annotated taxonomy of the international corpus of english

LANGUAGE	a system of human communication using structured arrangements of sounds or glyphs.
SPEECH	language expressed in sound.
WRITING	language recorded by symbols on a medium.
PRIVATE	writing for ourselves or other individuals presuming private information. Few copies.
PUBLIC	broadly disseminated writing for an unspecified readership assuming objective generally available knowledge about the world and the domain of discourse.
LITERAL	public writing employing the dictionary meanings of words.
INFORMATION	literal writing relating a sequence of events or the particular details of a thing.
INSTRUCTION	literal writing communicating a set of related concepts.
PERSUASION	literal writing developing an argument.
FIGURATIVE	public writing meant to be understood imaginatively.
PROSE	figurative writing using everyday speech rhythm and patterns.
POETRY	figurative writing expressing a heightened sensibility through image, rhyme, meter, sound and other aspects of style.

meeting them is not sufficient to establish a text as technical. It simply invites further attention. Additional features will be needed to distinguish technical from other kinds of texts.

3. Overview of the Study

We went about the task in this way. An ongoing *search of the literature* indicated that text typing lay at the intersection of a number of different fields of study—linguistics, psychology, cognitive and computer science. With two exceptions, we found little previous work of the sort we had embarked on. The exceptions are Douglas Biber's (1988) study of textual variation expressed in terms of a large number of linguistic features, and Karlgren and Cutting's (1994) application of Biber's methodology to categorize texts in the Brown corpus using a subset of his features. These two studies encouraged us to persevere in our approach to the task.

Next we *composed a list of features* which we deemed likely to characterize technical text. Our choices differ from Biber's in two ways. His are (a) objective counts and averages of lexical and syntactic characteristics; some of ours are objective as well, but others are subjective judgments. Biber's criteria are all (b) lexical and syntactic, while ours also assess a text's semantic, discourse and pragmatic aspects. And where both Biber and Karlgren and Cutting look at (c) typing text in general, we are only concerned with features that indicate text technicality. This task may be easier to accomplish because it is smaller. It does not require dealing with all the ways in which texts vary.

The items in our list were then evaluated by the team as a whole and a number were removed. The 42 remaining were then organized into an inventory (see Appendix A) and a procedure for rating text samples was developed. A rating manual was written, forms constructed, and a preliminary study performed on a single text sample to determine if the inventory appeared to have discriminant power. Results were positive. At that point the *main study* began. Nine quite varied text samples were selected and rated by eight members of the research team. To achieve a statistically significant sample size, an additional 13 raters were recruited by advertising in a local Usenet conference. These individuals rated one text on-line; those who returned a correctly filled-out assessment were mailed and returned forms and samples of the other eight texts. We deemed a rating satisfactory if it applied the methodology correctly and counted the right sort of instances for each item. The pertinence of any comments was also considered.

Two sorts of *analyses* were conducted on the data received. The rating procedure involved raters assessing each sample text in terms of each inventory item. In addition they assigned an

overall value to the text's technicality on whatever basis they chose. This overall technicality rating plays a key role in the study—measures of significance were calculated between it and each item value. Results were encouraging and have been acted on. Because the relationship between text features and overall technicality seems to be quite complex, multivariate analysis has been undertaken to explore it further, and in particular to determine to what degree the interaction between items would strengthen the overall discriminant power of the inventory. That analysis is still underway.

4. Related work

It has been observed that scientific researchers in many fields find it easier to work with technical text than with non-technical writing. As a result, a number of disciplines use the term in papers and consider the notion of types of text to some degree, and most of the papers in the literature review that follows make some reference to technicality. We had a number of issues to explore when we turned to existing research—the process of categorizing, the nature of text types, what typologies already exist—and many fields to investigate. Findings are therefore presented on a field-by-field basis; certain conclusions drawn from them have already been summarized in part in Section 2 and additional reference will be made in the discussion where appropriate.

4.1 Computer science

Foremost among the disciplines concerning themselves with technical text is computer science, where processing text is considered in the subfields of Natural Language Processing (NLP), Machine Translation (MT) and Information Retrieval (IR). In addition to constraining language by limiting the class of texts they deal with, NLP researchers alternatively use a *sublanguage* which is either particular to the domain (Grishman and Kittredge, 1986; Lehrberger, 1986), or which simplifies the vocabulary and grammar of the whole language (Kniffen *et al.*, 1989). The latter approach has found favor in multilingual commercial settings where technology is involved (Chervak *et al.*, 1996) and has been investigated as a subject in itself (Kittredge, 1982, 1987). Frail and Freedman (1991) articulate a set of related notions—concept predictability, 'the ability to predict the structure and sequence of the concepts defined for a domain of texts'; concept structural complexity, 'a high number and variety of marked and unmarked concept instances . . . [and] a large syntactic variety'; and grammaticality, on both the linguistic and orthographic level—which are significant characteristics of the fragmented sublanguage texts they seek to summarize by extracting instances of particular concepts.

As an inducement to us in our present endeavour, Kittredge (1987) reports that there is 'strong evidence that languages are more similar in sentence and text structure within scientific and technical writing than in nontechnical writing' and that 'corresponding sublanguages [of different languages] are often structurally more similar than two dissimilar sublanguages of the same language'. This suggests that technical writing has regularities which can be identified and assessed in a rating inventory.

In general NLP practitioners have skirted the issue, simply indicating that their subject texts are technical, or identifying documents that would generally be viewed as technical (Cordova, 1992; Boggess, 1993; Obermeier, 1985). Their attention lies elsewhere; in parsing or in some other problem in syntactic or semantic processing (Sutcliffe *et al.*, 1994a, b). In certain cases the domain in question is so limited—programming language statements—that the language

involved is in effect a quite restricted sublanguage (Rowe and Laitinen, 1995; Kontos and Cavouras, 1988).

Machine translation researchers have also faced the issue of typing text. Their endeavour is the most applied sort of language processing, and the variety of texts requiring translation in a modern organization stretch MT systems and force researchers to address differences between texts. Arnold (1989) acknowledges that although at present no typology exists, text type is crucial. He identifies five dimensions of text which can be restricted (thereby placing MT-suitable texts in the class of sublanguages):[4] semantic domain (subject field),[5] discourse type,[6] discourse structure, syntax and morphology, lexis. Sager (1982) remarks that translations are affected by the purpose to which they are put; this makes the MT task more 'committed' than that of the NLP generalist, who can ignore the purpose to which his construction is put and focus entirely on its veridicality.

A theme first enunciated by Sager in 1982 has regularly reappeared in the MT literature: for a MT system to be practical, it must be customized to the particulars of its text and the needs of its users. Roudaud (1992) is closest in intent to our purpose in this paper in his text classification effort undertaken in the context of the French GETA project. His group sought to characterize aeronautical maintenance manuals and job cards (procedural descriptions). The resulting description, 7 frequent and 8 infrequent phenomena, emphasizes grammar but includes lexical and semantic characteristics. This is similar to our own list and consistent with repertory grid analysis. Unfortunately, although Roudaud calls his work a typology, it is hardly that, describing a single sort of text.

Research has been undertaken into indexing machine-readable texts automatically. Vickery (1968) suggests that inclusion and exclusion lists frequently drive the process. An inclusion list contains likely descriptors (taken from the classification system), while an exclusion list enumerates the articles, propositions and abstract words (thing, theory, report) that should not appear.

Researchers in the IR field have taken an applied view of the subject. Paice and Jones (1993) present an approach integrating automatic abstracting and indexing of 'highly structured empirical research papers'. Gauch and Futrelle (1993) propose to exploit this structure in a document browser, while Lai and Manber (1991) suggest that hypertext systems would benefit from using it to provide users with a 'flying through hypertext' facility that would be akin to quickly skimming or flipping the pages of a book.

4.2 Linguistics

Our inability to find a substantial body of discussion about types of text or text typologies within traditional linguistics is both surprising and disconcerting.

In a 1977 paper, Hutchins identifies two relatively distinct approaches in linguistics to the analysis of text structure focusing on *microstructure* and *macrostructure* respectively. The former is the level of anaphora, ellipsis and reference, etc., while the latter deals with how episodes develop from one another and how paragraphs and chapters are built into a 'cohesive whole'. Hutchins considers the scientific paper, defined as an 'article of a learned journal which argues for the revision of some accepted opinion in some area of academic study'. In such papers the structural form of the text 'is determined by the particular communicational functions it has to serve'. He supports this opinion by a Cook's tour of mainstream linguistic analytics applied to text microstructure or macrostructure: thematic progression, semantic progression,[7] discourse mode[8] and episodic function.[9] Wenger (1993) discerns similar lower-

level and higher-level structure in technical text, finding that the higher level has a greater effect on reader performance.

Hutchins' subject matter is conceived more narrowly than technical text. His classification of scientific papers into controlled-experiment, hypothesis-testing, or technique-description is inapplicable to repair manuals or government documents, to say nothing of fragments like labels or assembly instructions. Nevertheless, he makes a convincing argument for an overarching structure in all *coherent* texts akin to thesis-antithesis-synthesis, a structure which expresses our subconscious expectation to be told a good story.

In his 1995 doctoral dissertation, M. N.Zafio investigates three classes of texts (scientific, technical and popular) that talk about technology in five domains (electricity, electronics, aerospace, nuclear and mechanics). He considers 45 short French tests drawn equally from the categories produced by intersecting these classes and domains. N.Zafio looks at textual characteristics on the five levels of *enunciative* (linguistic indications signaling the presence of the narrator or the reader), *lexical*, *syntactic*, *macro-structural* (clauses and the type of information they convey), and *co-textual* (non-linguistics signs such as figures, drawings, photographs, formulae, etc.). Main component analysis shows that the three text classes are in fact distinct, with the scientific and technical classes resembling each other syntactically in their use of domain-specific terms, numerous subordinate clauses, explanations, references and inanimates.

Douglas Biber has been engaged in a comprehensive and sustained investigation of text typology for more than 10 years. In a 1988 paper he reports the general view within linguistics that 'written language is structurally elaborated, complex, formal, and abstract, while spoken language is concrete, context-dependent, and structurally simple' (p. 5). He then identifies 67 linguistic features upon which to classify text (see Table 2). Six dimensional scales are 'determined on the basis of a consistent co-occurrence pattern among features' (p. 13), which are for the most part unambiguous linguistic phenomena such as passives, nominalizations, contractions, etc. Underlying relations are defined in terms of these dimensions and 'specify the ways in which any two genres are linguistically similar and the extent to which they are similar' (p. 55).

Table 2. Lexical and syntactic features used in Biber's classification

1st person pronouns	general hedges	private verbs
2nd person pronouns	indefinite pronouns	present tense verbs
3rd person pronouns	infinitives	pronoun IT
adverbs	necessity modals	public verbs
agentless passives	nominalizations	SEEM/APPEAR
amplifiers	non-phrasal coordination	sentence relatives
analytic negation	nouns	split auxiliaries
attributive adjectives	other adverbial subordinators	sausive verbs
BE as main verb	past participial WHZ deletions	synthetic negation
BY-passives	past participial clauses	THAT clauses as adjective complements
causative subordination	past tense verbs	THAT clauses as verb complements
concessive subordination	perfect aspect verbs	THAT deletion
conditional subordination	pied piping constructions	THAT relative clause on object positions
conjuncts	phrasal coordination	time adverbials
contractions	place adverbials	type/token ratio
demonstrative pronouns	possibility modals	WH clauses
discourse particles	predicative adjectives	WH questions
downtoners	prediction modals	WH relative clauses on object positions
existential BE	prepositions	WH relative clauses on subject positions
final prepositions	present participial clauses	word length
general emphatics	present participial WHZ deletions	

Experiments were run on 2000-word samples from each of the 19 genres recognized in the London-Lund and LOB corpora. These include newspapers, broadcasts, official documents, academic prose, biographies, conversations and fiction. For purposes of analysis the 67 linguistic features in Table 2 were grouped positively and negatively into seven (unnamed) factors. The five dimensions[10] identified by analysis of these factors are: involved vs informational production, narrative vs non-narrative concerns, explicit vs situation-dependent reference, overt expression of persuasion and abstract vs non-abstract information. Many of the dimensions and text types seem applicable to technical text but not exclusively so. More important for our purposes is the fact that they are equally abstract, and therefore do not advance our goal of identifying concrete features which indicate text technicality. Biber subsequently identified eight types in these two corpora[11] (1989), clusters of texts which are 'markedly similar to one another with respect to their dimensional characteristics': intimate personal interaction, informational interaction, scientific exposition, learned exposition, imaginative narrative, general narrative exposition, situated reportage and involved persuasion. Technical text could be expected to span at least two of the expository types, scientific and learned, and perhaps more; and a look at the genre making up each cluster suggests that, for most types, certain of the texts would generally be considered technical and others not. Biber's more recent work applies his analysis to new corpora critically (1993) and extends it to other languages (1995).

Biber classified texts in the LOB and London-Lund corpora. In 1994 Karlgren and Cutting applied his technique to the Brown corpus, the third corpus in general use in the 1980s and early 1990s. These authors concentrated on automatability, selecting only features which are easy to compute in a mechanical way for a given text. Biber employed factor analysis to classify texts into the most distinct categories possible; Karlgren and Cutting constructed functions to discriminate among the texts in the existing Brown corpus categories, grouped into two or four categories or taken as-is. They found that discriminant functions were successful classifiers in inverse relationship to the number of categories involved. 'Error rates climb steeply with the number of categories tested for', they note, suggesting that content-based classification be used together with filtering based on lower-level cues.

In a review (1995) of Biber (1995) Adam Kilgarriff discusses that author's research program as a whole. He deems it important for practical reasons: natural language contains many syntactic constructions and anomalies and designers of automated language processing systems must determine which ones are likely to occur in application text. Kilgarriff identifies relevant work in computational linguistics (sublanguages) and in sociolinguistics, but notes they lack the methodology Biber offers, which he summarizes as:

1. Gather a set of text samples to cover a wide range of language varieties;
2. Enter them into the computer;
3. Identify a set of linguistic features likely to serve as discriminators for different varieties;
4. Count the number of occurrences of each linguistic feature in each text sample;
5. Perform a factor analysis to identify which linguistic features tend to co-occur in texts. Members of the set of factors or *dimensions* each carry a weighting for each feature;
6. Interpret each dimension to identify what linguistic features and communicative functions high-positive and high-negative values on the dimension correspond to.

Kilgarriff judges these steps to 'resoundingly meet' Biber's five desiderata of a typing methodology. These are that it be *corpus-based*, consider the *full range* of language register, look at a *wide variety* of features, address *interactions* between them without preconceptions,

and involve both *quantitative and qualitative* methods (for respectively identifying and interpreting patterns of variation). Notwithstanding the need for such a technique and the efficacy of the one in question, Kilgarriff remarks that 'as yet the methodology has only been used by Biber and a small group of collaborators'. He believes this is because it is 'technically difficult and time-consuming to implement' but further notes that 'this is the only such machinery there is'.

4.3. Other fields

What conclusions can we draw from researchers in other fields? Very little, apparently—the strongest impression we received was of a paucity of relevant work elsewhere. For instance, although it seems to us likely that philosophers, semanticists, literary analysts and investigators of media and mass communications might have something to say about types of text and their structure, after fairly diligent search we did not find anything noteworthy in these disciplines. Content analysis, which concerns itself with the content of communication, considers text typing only in a very high-level and abstract way. Psychology's focus is on the situational or personal dimensions of writing—the development and exercise of writing skills. Psychologists must however deal with experimental situations similar to ours, and their repertory grid analysis data acquisition technique (Dillon and McKnight, 1990) is a more robust if less rigorous alternative to the traditional statistical analysis employed here. Psychologists' interest in the writing context means they have things to say about a text's purpose and pragmatics (Cooper and Matsuhashi, 1983; Britton and Graessner, 1975). Finally, libraries provide the preeminent classification of texts in the real world; and library science concerns itself with the theory of categorization and its practice in indexing (Daily, 1971, 1976). The manifest inadequacy of the Dewey Decimal or Library of Congress classification systems for our purposes drives home how minor a factor a text's topic is in determining its technicality.

5. Developing the inventory

5.1. Enumerating items

With some idea of how others view the matter, we began to identify textual features which seemed to be likely indicators of text technicality. Candidates came from a variety of sources—from among those mentioned by authors discussed in the last section, from introspection and self-observation, and from inspecting texts, comparing them to one another and analyzing differences. At this point our objective was to be encyclopedic, so no restriction was placed on the kinds of feature envisioned and no consideration was given to how easy it would be in practice to assess a property.

Each of us worked independently. Candidate items were presented at a series of weekly meetings where each was discussed and evaluated. This scrutiny accumulated a good deal of information for each feature that passed muster; in addition to characterizing the property itself, the vetting process also produced a rationale of why it might mark technicality and, where this made sense, examples of its occurrence in text. Much of this information was reproduced in the materials given to the raters who participated in the main study. Of all the features considered, 42 seemed likely enough positive or negative indicators of technicality to warrant testing on actual texts. These appear in Table 3. A description of what each means can be found in the Appendix.

Table 3. Text technicality inventory items

1	introduction or a summary setting out the content	22	colloqualisms
2	table of contents, index, appendices	23	vague terms
3	citations	24	meaning conveyed by denotation
4	quoted dialog	25	explicit analogs only
5	special fonts and punctuation	26	figurative language
6	title and headings follow a convention	27	examples
7	text clearly identifies its topic	28	ellipsis
8	text focused on its topic	29	text binders, general hedges
9	communicates a body of knowledge	30	discourse particles, general emphatics
10	serious and objective treatment	31	use of the third person
11	orderly and logical presentation of ideas	32	causative subordination
12	discussion becomes more complex as it proceeeds	33	complex sentence structure
13	material organized hierarchically	34	domain-specific verbs with single senses
14	topics accessible at random	35	verbs in present tense
15	writing grammatical & sentences well-formed	36	stative sentences
16	terms defined in the text	37	passive constructions
17	non-textual elements	38	interrogative and imperative sentences
18	itemized lists	39	statements of opinion or belief
19	domain-specific terminology	40	generic reference
20	conventional expressions or formulae	41	unambiguous references
21	humor, sarcasm, invective	42	nominalizations

These properties can be organized in a variety of ways. Two seem pertinent here. One scheme is based on the traditional spheres of linguistics: *lexical, syntactic, semantic* and *pragmatic*. Taken together, these four categories comprise a hierarchy of scoping that moves from a focus on the individual word, to the broad expanse of the shared situational context within which both author and reader interpret a text's content and presentation.

An alternative point of view groups text features according to the size of the textual item they apply to: single *words*, entire *sentences*, *sections*, or the whole *document*. Table 4 describes these two sets of criteria more exactly.

Taken together, these two viewpoints define Table 5, a table into whose cells, items in the inventory can all be classified. An item in the uppermost left-hand cell involves the morphology of individual words; an item in the cell two rows down in the same column concerns their

Table 4. Major dimensions of properties characterizing texts

Word	words, collocations, and non-lexical elements such as diagrams or graphics
Sentence	phrases and clauses
Section	text structure above the sentence: paragraphs, sections, chapters
Document	the entire text taken as a whole
Lexical	the individual words and fixed expressions
Syntactic	the grammatical constructions built of lexical elements
Semantic	the meanings of lexical and syntactic elements, and larger discourse structures
Pragmatic	the non-linguistic context within which the text is to be interpreted

Table 5. Classification of properties in the inventory

	Word	Sentence	Section	Document
Lexical	17, 19, 22	20		3, 5
Syntactic	29, 30, 42	15, 31, 33, 35, 36, 37, 38	4, 18, 28	1, 2, 13
Semantic	16, 23, 24, 34, 40	32, 41		25, 26, 27
Pragmatic		39	6, 8, 14	7, 9, 10, 11, 12, 21

meaning. An item in the lowermost right-hand cell deals with factors, especially organization, that apply to the whole document.

This grid helped us organize the rating form. Assignment of items to a single cell was driven by the need to provide raters with instructions and a procedure that are as clear and unambiguous as possible. Assignments to cells are arguable, and the table reflects majority opinion.

5.2. Developing a rating methodology

Simply listing and evaluating items for the inventory made us more aware of the various textual features that mark technicality. This was an entirely subjective experience however, and a more formal methodology was needed to test the inventory items. The first step in developing it was to establish *what class of evaluation* was possible. The inexact nature of both technicality and of many of the candidate inventory features put objective, empirical measurement out of reach and required that trust be placed in the impressions of human raters. Statistical theory indicates that when multiple raters are used, these judgments are verdical within tolerable confidence limits if the number of raters is large enough and if their ratings tend to agree. The typical assessment in the inventory, then, is the rater's estimate of a property's occurrence in a sample text.

The next step was to determine *what kind of judgment* was appropriate to each property. A property can be quantitative, in which case counts of instances are in order; or it can be qualitative, where values are binary or graded. A binary property is either present or absent—does the document have a Table of Contents or not? A graded property is harder to estimate; it is usually present to some degree. To help determine this degree, raters were asked to count instances of the given property in the sample text, to compare its occurrence there with its occurrence in writing in general, and then to record on a Likert scale whether they deemed it to be much more, more, about as, less, or much less common. The anchors of this five-point scale were PRESENT and ABSENT. An example of a graded property is passive constructions—in the rater's impression, does the text have fewer or more such constructions than writing in general?

In an ideal world raters would have the time and motivation to inspect word instances in texts of any size. In reality this is not possible. The limit of human endurance kept us from using many of the low-level textual features that Biber used in his set of parameters; we simply could not expect raters to compute average sentence lengths etc. with any degree of accuracy, or even at all. Such tasks are best suited to automation. The inability to use textual characteristics on this lowest level did not seem to be an important setback to the development of the inventory; few of them had figured in our enumeration of plausible technicality indicators. Table 5 indicates however that a large number of features have WORD scope. A *sampling* regimen was put in place for features on the sentence, paragraph and page[12] levels—word instances would be sought in sentence samples, sentence instances in paragraph samples, and paragraph instances in page samples or the entire document. Rating becomes much less onerous when raters can look in just a sample rather than the whole text for the instances on which to base their assessment of a property.

Sampling was performed in this way. Raters were instructed to pick a sample of any unit, sentence, paragraph or page, from a text which contained more than 10 instances of it. Samples on each level were picked from the next level higher sample or from the entire document if no samples existed. Thus sentences would be chosen from sample paragraphs; failing that, from sample pages or the whole document. Ten sentences would therefore be picked from a text

composed of 17 sentences in four paragraphs while all four paragraphs would be used. A 23-page text would be sampled on all three levels—page, paragraph and sentence. Raters were directed to choose samples as randomly as possible while ensuring that they were representative of the text. Random sampling is appropriate when the population from which the sample is drawn is homogeneous. This is not the case with the class of documents we are interested in; researchers in many fields have found technical documents to be in fact highly structured (Paice and Jones, 1993; Gauch and Futrelle, 1993). Accordingly, raters were asked to exclude items not representative of the main body of the text from samples. Picking each sample from the members of the next level higher sample tends to ensure representativeness as well because this selection is already qualified as representative.

The introduction of sampling required that *the scope of each inventory item* be determined. This categorization generally follows that presented in Table 5, with items in the SECTION column divided between page and paragraph sampling. At this point a practical inventory was beginning to emerge; to make it easiest to fill out, items were grouped by scope from greatest to least so that raters can keep focused on the same set of samples until all items involving it are completed. This should help them perform their task in an orderly and methodical way.

To further encourage this to happen, raters were asked to *count instances* and provide examples for almost all items on the page, paragraph and sentence levels. In longer texts these are sampled, graded items; the hardest ones to rate. This data will permit a secondary analysis to be performed on a subset of the non-sampled data whereby we will be able to assess raters' skill at identifying the underlying syntactic phenomena on which they based their assessment of particular items. It will also clarify whether people tend to share the same understanding of what a passive sentence, a noun phrase, or a 'be' verb are.

The final step was perhaps the most important. A 43rd item was added to the inventory—the rater's graded estimate of a text's *overall technicality* based on whatever grounds he or she found meaningful. This addition provided a value against which to test individual features in pairwise correlations and other statistical procedures.

The rating form in its final format appears in Appendix A. The manner in which its items are presented is the result of a process of gradual refinement which took place over a number of months in response to the comments of early users. A companion manual issued to raters is not reproduced here, but is available from the authors on request. This manual explained the purpose of the experiment and the rating procedure, and told when and how to pick a sample. It reproduces entries for the items on the rating form, supplemented with additional details and the rationale behind our selection of each.

6. Conduct of the study

6.1. The preliminary study

To try out the newly-developed rating materials, six members of the Knowledge Acquisition and Machine Learning (KAML) laboratory at the University of Ottawa Department of Computer Science rated a chapter from an introductory text on accounting, a document deemed likely to be quite technical. Results of this preliminary study indicated that the inventory as constituted offered some prospect of assessing text technicality: (a) raters' estimates of various items *tended to agree* with one another; (b) certain properties show a *high positive* or *negative correlation* with technicality. Overall technicality of the accounting text chapter was not assessed in this administration but was simply assumed to be quite high.

Raters were voluble in their comments on the process. All agreed it was quite time-consuming, an observation which was made repeatedly over the course of the study. Perhaps this demand for a good deal of effort on their part motivated raters to examine the process. Suggestions for simplifications and improvements were reviewed in a post-mortem and many were put in place for the main study that followed. The modifications proposed dealt mostly with the rating methodology and not with the content of the inventory. Low-level quantitative items computing the average word and sentence length in the text were dropped from it as too time-consuming, and a graded measure of the frequency of nominalizations was added as was a final overall rating of technicality. One item was renamed to facilitate rating (from *stative sentences*[13] to *'be' sentences*), and another was restated in an inverted fashion (from *ambiguous references* to *unambiguous references*) when exceptions proved easier to spot than conforming instances. In general the inventory has remained much as it was initially framed.

6.2. The main study

With encouraging results from the preliminary study, it was decided to embark on an administration of the inventory in sufficient numbers to produce meaningful statistics. It was further determined that a number of texts of varying degrees of technicality would be rated: we had reason to believe that certain of the inventory items measured technicality; but would they correctly measure the *absence* of technicality in non-technical texts? Nine texts were chosen to be used in the main study for this purpose. Selection was necessarily ad hoc, as the study itself was an attempt to develop a tool to assess the feature in question. Results show that the chosen texts did in fact vary widely in overall technicality.

The chosen texts were:

• a poem;	24 li.	poem	P
• a commercial brochure on lawn care;	20 para.	brochure	B
• an act from a play;	10 pp.	play act	PA
• an instruction sheet for medication;	12 para.	label	L
• a retail merchandise catalogue entry;	16 li.	catalogue entry	CE
• a recipe from a cookbook;	33 li.	recipe	R
• an academic paper;	11 pp.	academic paper	AP
• a literary essay;	19 pp.	literary essay	LE
• a short story.	6 pp.	story	S

The right-hand columns above list the length, short name and abbreviation for each text used in the discussion that follows. Details about each text are given in the Analysis section where their associated technicality rating is discussed.

Eight members of the KAML lab team rated all of these texts. Two others rated one or more. Given a potential 42 independent variables (the number of rating items) and the length and difficulty of the task, it was decided to recruit additional raters to ensure that more than sufficient data was on hand to provide correlation results at the 0.05 confidence level.

To this end an ad was placed on a local Usenet conference offering a small stipend to individuals who completed satisfactory ratings for the nine sample texts. Respondents received ASCII copies of the rating materials—form, manual, instructions—by email, along with one text (the short story) which had no graphic or format styling and therefore would not suffer by being rendered as straight text. Recipients who returned an email rating of this text that demonstrated a grasp of English grammar and an adequate attention to detail were offered the

Table 6. Participation rates of Usenet group

	Initial response	Participate?	Returned email rating	Received mailed rating	Returned mailed rating
First posting	27	22	11	7	5
Second posting	22	21	1	1	1
Third posting	17	14	4	4	4
Total	66	57	16	12	10

opportunity of rating the additional eight texts. Those who accepted were sent the texts by surface mail. Most returned completed rating forms.[14]

Three postings of the Usenet ad over the winter of 1995 and spring of 1996 elicited responses from 10 raters who completed the task, which was estimated to require approximately 10 hours. An additional four raters were recruited by word of mouth. Given the effort involved, no deadlines were imposed on these people: we wanted the best work possible. This is largely responsible for the length of the study. Table 6 summarizes participation.

Although a good deal of effort was spent trying to make the rating materials easy to follow, errors did occur. Perhaps this is due to the inherent complexity of the task, but we would make further refinements to future versions of the rating materials, both form and manual. The greatest problems here were with selecting the right text and using the Likert scale correctly. The first problem arose when, in order to give the context for the text being rated, the sample provided was the entire page (catalogue) or two-page spread (recipes). Some raters assessed *all* recipes or catalogue items on the page rather than just the one indicated. There were two problems with scale items. A Likert scale is composed of a number of cells in a line meant to express the range of possible values between anchor terms at either end. Some raters circled an anchor instead of ticking a cell and one rater ticked more than one cell. A perennial problem in a lengthy questionnaire is items being skipped; that happened here, and was further confounded by our direction to simply ignore items that did not apply to the given text. A better instruction would have been to use 'N/A'. All data were entered and evaluated for use in the analysis. The rating in which multiple points on the scale were checked for a number of items was dropped from consideration.

7. Analysis

7.1. Raw data

Table 7 summarizes the results of rating. It presents the average rating assigned by raters to each item in the inventory for each of the nine sample texts whose abbreviations head its columns. See Section 6.2 for a key to these abbreviations. Averages for binary items have been normalized over the 1:5 range of scaled items. Averages for the three items which were only counted are reported unchanged, separated from the rest of the table by dotted lines.

After compiling this data we began analysis. We first tested whether items correlated significantly with technicality on an individual basis. This was accomplished using the CORREL function in Excel 5.0 for the 39 items which were deemed to be normally distributed (items #1 through #15, #19 through #42). Contingency table analysis in GenStat 5 (release 3.2) was also performed on all 42 items, which fall into either a normal or a Poisson distribution. The output of the latter computation was evaluated for significance at the 0.05 level using values χ^2 values (df:4, 9.49; df:16, 26.3) and f critical values (df:24/~155, 1.60) as appropriate.

Table 7. Average values for inventory items, by text

#	Item	PA	S	P	LE	B	CE	R	L	AP
1	introduction	1.0	2.3	1.0	2.3	4.0	2.1	2.6	2.6	4.7
2	ToC, Index	1.0	1.0	1.0	1.4	1.4	2.1	1.5	1.6	2.0
3	citations	1.2	1.2	1.0	2.8	1.4	1.0	1.2	1.4	5.0
4	quoted dialog	4.8	4.0	1.0	2.8	1.2	1.0	1.0	1.0	2.6
5	fonts & punctuation	4.3	2.4	1.6	3.5	3.5	4.5	3.4	5.0	5.0
6	titles, headings	2.1	1.3	1.6	1.7	3.9	3.3	3.2	3.7	5.0
7	topic identified	1.5	2.3	2.1	4.8	4.8	4.8	4.8	4.8	5.0
8	focus on topic	1.4	3.2	3.1	4.6	5.0	5.0	5.0	5.0	4.8
9	communicate knowledge	1.2	2.0	2.1	4.8	5.0	4.6	4.8	4.3	5.0
10	serious, objective	1.2	1.7	1.2	4.0	5.0	4.8	5.0	4.5	5.0
11	orderly & logical	1.2	2.9	1.8	3.7	5.0	4.3	4.6	4.2	5.0
12	increase complex	1.4	1.8	1.4	2.6	2.7	1.0	1.7	1.2	4.8
13	hierarchic organization	1.0	1.2	1.0	1.5	4.6	2.1	3.2	3.4	4.8
14	random access	1.2	1.0	1.2	1.9	4.6	3.4	3.5	4.3	4.0
15	grammatical	2.3	4.0	2.7	4.5	4.0	1.7	3.9	3.1	4.8
16	defined terms	0.2	0.5	0.1	0.3	1.4	0.3	0.0	1.1	8.4
17	non-textual elements	2.8	0.2	0.0	0.1	1.3	3.1	0.2	8.9	10.0
18	lists	0.0	0.8	0.0	1.1	3.7	2.2	3.3	1.6	3.2
19	terminology	1.6	1.7	1.0	3.3	3.4	4.3	3.5	4.3	4.7
20	conventions	1.4	1.2	1.2	1.1	1.5	1.5	1.9	1.5	3.9
21	humor, invective	4.0	3.4	1.3	1.1	1.1	1.0	1.4	1.0	1.0
22	colloquialisms	4.2	3.2	1.5	1.4	1.4	1.1	1.4	1.0	1.0
23	vague terms	2.3	2.1	2.0	2.0	1.5	1.3	1.3	1.2	1.7
24	denotation	2.4	3.0	2.9	2.7	1.9	2.4	2.4	2.2	3.1
25	explicit analogs	1.3	2.0	2.2	2.0	1.3	1.0	1.2	1.3	1.6
26	figurative language	2.6	3.4	3.8	3.0	2.1	1.0	1.4	1.0	1.2
27	examples	1.0	1.3	1.0	1.4	1.6	1.2	1.1	1.4	3.4
28	ellipsis	2.8	2.6	2.7	2.8	2.2	3.1	2.3	2.7	1.6
29	binders & hedges	1.7	2.5	1.9	2.4	2.7	1.2	1.6	1.5	3.4
30	particles, emphatics	2.5	2.6	1.6	2.2	1.8	1.0	1.2	1.0	1.9
31	third person	3.3	3.9	3.4	4.1	2.4	1.4	2.3	1.6	3.7
32	causative subordination	1.3	2.4	1.9	2.8	2.0	1.0	1.7	3.5	2.5
33	complex sentences	2.2	3.7	3.6	4.7	2.7	1.3	2.2	3.5	4.1
34	same-sense domain verbs	1.2	1.5	1.4	1.6	2.5	1.5	2.1	1.6	2.6
35	present tense	4.3	2.6	3.6	4.0	4.1	4.5	3.8	4.5	4.1
36	stative sentences	2.9	2.8	1.9	3.3	2.9	1.2	1.3	1.8	2.9
37	passives	1.5	1.9	1.8	2.0	2.9	1.2	1.5	1.5	3.0
38	interrogatives & imperatives	3.5	2.7	1.6	1.1	3.3	1.2	3.3	3.8	1.5
39	opinions & belief	2.1	2.4	2.3	3.2	2.1	1.2	1.6	1.5	2.0
40	generic reference	1.9	2.2	2.9	2.4	3.6	2.8	2.6	3.4	3.2
41	unambiguous refs	1.8	2.5	2.2	2.2	2.5	2.4	2.4	2.8	2.0
42	nominalizations	1.3	1.9	2.1	1.7	2.5	2.4	1.2	2.3	2.5
43	technicality	1.2	1.3	1.3	2.8	3.6	3.9	3.9	4.0	5.0

These two procedures differ in the way they view the technicality data. Correlation treats values for the nine sample texts as constituting a single arithmetic set, while contingency table analysis sees each value as representing a unique category in an enumerated set. The latter procedure postulates as many orthogonal variates, hypothesized independent variables, as are required to explain the data; in this case, two variates account for enough of the total variance to suggest that they are significant. Results of these two analyses appear in columns 3 through 6 of Table 8. Columns 7 and 8 present initial results from a multivariate analysis; they are discussed below.

Considering how individual items vary with technicality in isolation serves as a first approximation to analysis. The two procedures used for this purpose differ to some degree as to

Table 8. Descriptive statistics for inventory items, by text

#	Item	Corr	Signif?	X^2/F	Signif?	Vect I	Vect II
1	introduction	0.79	TRUE	8.65	FALSE	-0.587	-0.539
2	ToC, Index	0.90	TRUE	6.07	FALSE	0.265	-1.389
3	citations	0.52	FALSE	15.39	TRUE	0.707	0.496
4	quoted dialog	-0.54	FALSE	38.56	TRUE	-0.326	0.094
5	fonts & punctuation	0.69	FALSE	12.92	TRUE	-0.348	-0.532
6	titles, headings	0.91	TRUE	26.06	TRUE	0.256	0.459
7	topic identified	0.93	TRUE	46.06	TRUE	2.074	-0.623
8	focus on topic	0.85	TRUE	44.75	TRUE	-0.244	-0.086
9	communicate knowledge	0.91	TRUE	38.91	TRUE	-0.287	-0.304
10	serious, objective	0.95	TRUE	69.21	TRUE	1.401	-0.470
11	orderly & logical	0.91	TRUE	30.45	TRUE	0.292	0.306
12	increase complex	0.48	FALSE	7.78	FALSE	0.866	0.167
13	hierarchic organization	0.86	TRUE	16.27	TRUE	-0.353	1.096
14	random access	0.91	TRUE	41.15	TRUE	0.298	-0.696
15	grammatical	0.32	FALSE	8.44	FALSE	0.108	-0.535
16	defined terms	—	—	14.00	TRUE		
17	non-textual elements	—	—	7.12	TRUE		
18	lists	—	—	2.12	TRUE		
19	terminology	0.97	TRUE	69.52	TRUE	-0.889	-0.259
20	conventions	0.66	FALSE	47.11	TRUE	-0.990	-1.489
21	humor, invective	-0.73	TRUE	32.95	TRUE	-0.159	-0.219
22	colloquialisms	-0.75	TRUE	51.85	TRUE	0.440	0.185
23	vague terms	-0.79	TRUE	17.84	FALSE	-0.214	-0.661
24	denotation	-0.20	FALSE	21.93	FALSE	0.311	-0.005
25	explicit analogs	-0.54	FALSE	33.69	TRUE	-0.631	-0.311
26	figurative language	-0.88	TRUE	24.74	FALSE	-0.080	0.753
27	examples	0.62	FALSE	19.72	FALSE	0.116	-0.301
28	ellipsis	-0.52	FALSE	43.32	TRUE	-0.272	0.876
29	binders & hedges	0.18	FALSE	36.14	TRUE	0.169	1.061
30	particles, emphatics	-0.61	FALSE	31.65	TRUE	0.840	-0.374
31	third person	-0.48	FALSE	27.63	TRUE	-0.047	-0.560
32	causative subordination	0.22	FALSE	30.77	TRUE	0.060	-0.881
33	complex sentences	-0.06	FALSE	14.21	FALSE	-0.240	0.314
34	same-sense domain verbs	0.73	TRUE	72.04	TRUE	-0.929	-2.055
35	present tense	0.53	FALSE	59.84	TRUE	-1.006	-0.018
36	stative sentences	-0.26	FALSE	57.44	TRUE	0.216	0.012
37	passives	0.33	FALSE	24.56	FALSE	0.598	-0.201
38	interrogatives & imperatives	-0.07	FALSE	47.11	TRUE	-0.459	0.733
39	opinions & belief	-0.48	FALSE	12.26	FALSE	-0.867	-0.453
40	generic reference	0.67	FALSE	34.44	TRUE	-0.316	0.610
41	unambiguous refs	0.29	FALSE	39.76	TRUE	-0.516	-0.276
42	nominalizations	0.46	FALSE	38.79	TRUE	-0.060	0.291

which items they indicate to be significant predictors; while noting which items were identified by both procedures, we have used the union of the two sets to revise the inventory. Moreover, certain features which most observers would deem strong indicators of technicality (such as #3, *citations in the text*) appear here correlated with it only weakly. These facts suggest that the technicality of a text is a complex matter—that the data must be interpreted thoughtfully, and that further analysis on a multivariate basis is appropriate. A canonical variate analysis of the data is in fact currently being explored. Canonical variate analysis is a type of discriminant analysis which makes no assumptions about the distribution of individuals from their occurrence in the sample. This class of procedures postulates as many orthogonal variates as are required to account for the data. Its initial results suggest that two variates, hypothesized variables, are significantly involved in the data as currently formulated. Vector coordinate

values for these two variates, column 7 and 8 in Table 8, are reproduced to give an indication of the manner in which variance in the dataset is shared amongst the items in the inventory.

7.2. Discussion

Because the result of multivariate analyse is still uncertain, discussion will be limited to what has been learned about individual items. In addition, given the differences between the χ^2 and correlation results for individual items, the strength of the relationship between an item and technicality will be described in general rather than particular terms: 'strongly indicative' rather than 'correlated at the 0.97 level'.

Before any inferences can be drawn, the data must first pass a qualifying test: do the sample texts vary on the key measure, technicality? This requirement must be satisfied before we can determine whether any of the 42 features in the inventory are good predictors of that feature. The ratings for technicality in Table 7 (#43) show that raters did indeed deem the chosen texts to vary widely on this feature and, further, that their values for it span almost the entire range of the scale. This suggests that the set of sample texts offers at least the potential of providing indices of varying degrees of technicality. The next questions arise naturally: do any of the 42 candidate predictors actually track the key feature, and if so, which ones and to what degree? The answer here is also 'yes', though a more qualified one.

But let us first look at raters' assessments of overall document technicality. For convenience the relevant row from Table 7 is duplicated below, supplemented with figures for each text giving the standard deviation for the technicality rating and the average standard deviation for normalized values of all normally-distributed rating items.

The measures of deviation first. Average figures are fairly consistent across the nine texts, ranging from 1.01 to 1.46. Superficially, the narrow range suggests that raters tended to treat all texts similarly; the value itself, that raters tended to agree on ratings to a certain degree. Given a hypothetical average item rating of 2.5, 68% of raters picked a value between 1.04 and 3.96 for the literary essay (LE) and between 1.49 and 3.51 for the academic paper (AP). Values can range from 1 to 5, so the standard deviations are fairly great. However in every case save that of the catalogue entry (CE), deviation in the rating of technicality itself is lower, sometimes much lower, than the average. People tended to agree on the overall rating of a document's technicality.

These technicality ratings show the nine documents fall into four groups. Three are quite lacking in technicality: an *act* from the *play Who's Afraid of Virginia Woolf?* which is almost all colloquial dialogue, the *poem The Snake* by Emily Dickinson, and the O. Henry *short story The Gift of the Magi*. All three are fiction. The *literary essay*, an excerpt of Chapter VIII from Henry James' *Mont Ste. Michel and Chartres*, falls in a class by itself—tending towards technicality, but rather less than more. Though descriptive of something in the real world (the stained glass windows in Chartres cathedral), James' language is florid and overwrought by contemporary standards. Four documents constitute a third group. These texts have nothing to

Table 9. Overall technicality statistics, by text

#	Item	PA	S	P	LE	B	CE	R	L	AP
43	Technicality rating	1.2	1.3	1.3	2.8	3.6	3.9	3.9	4.0	5.0
	Technicality standard deviation	0.53	0.54	0.64	1.03	1.02	1.15	0.99	1.10	0.21
	Average standard deviation	1.15	1.44	1.16	1.46	1.24	1.11	1.29	1.25	1.01

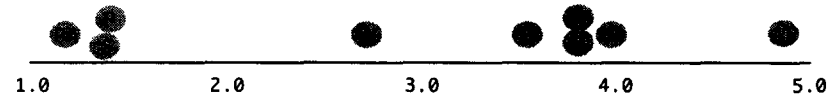

Fig. 2. A graph of the sample texts' technicality.

do with entertainment and represent the great body of written material people use in everyday life. They include a *brochure* published by a seed company telling how to grow a lawn, and *entry* from a retailer's *catalogue*, a *recipe* from Irma Rombauer's *The Joy of Cooking*, and the dosology sheet for a cold medicine (entitled somewhat misleadingly *label*). An *academic paper* (Philip Resnik: 'WordNet and Distributional Analysis: A Class-based Approach to Lexical Discovery') stands by itself at the top of the scale. Every rater bar one judged this text to be wholly technical.

The technicality dimension of the nine sample texts has been graphed in Figure 2. The values assigned to the various texts seem quite reasonable to us, as does the ordering that results; and for seven of the nine texts, the variance in raters' estimates of technicality is lower than the average variance for all features, suggesting that people tend to agree with one another in their assessment of technicality.

The 42 items that comprise the body of the inventory reflect these technicality values to some degree, the majority at some level of significance. Twelve of the items exceed the χ^2 critical value and correlate with technicality at the 0.70 level or higher. Others were found to be good predictors by a single procedure: 15 more items exceed the χ^2 value, and another four correlate significantly with technicality. All in all, 32 of the 42 items in the inventory vary meaningfully with the target feature according to one or the other analysis or both. These items have been retained in the revised inventory.

The following narrative summarizes features that were found to be meaningful predictors of text technicality. Individual items are italicized. Items which both passed the χ^2 test and possess significant correlation coefficients are listed in the first paragraph. Others which were shown to be meaningful by a single procedure follow after.

> A technical text is quite likely to have an *identified topic* on which it is *focused*. It will *communicate knowledge* about that topic in a *serious and objective* manner, developing its thought in a *logical, orderly* way. This produces a document with a *hierarchical organization* in which information can be *accessed at random*. Sections in such documents often bear *titles or headings*. Authors typically use *terminology* specific to the domain and avoid *colloquialisms* and *humor or invective*. Their writing is likely to use *same-sense domain verbs* and to make *generic references* (talk about classes rather than individuals).
>
> Technical texts frequently have an *introduction* and a *table of contents or index*. They may use *citations* but avoid *quoted dialogue*. Material may be presented with *special fonts or punctuation* or according to some commonly-understood *convention*. It avoids *vague terms* or *figurative language* and tends to use *explicit analogies, unambiguous references* and *nominalizations*.
>
> A number of syntactic characteristics suggest technicality. Technical writing uses few *interrogative or imperative* sentences, but sentences incorporating some form of '*be*' are common. Statements are often couched in the *third person* and the *present tense* and employ *subordination suggesting cause* and effect. *Binders and hedges* are used to knit the narrative together. Writing can be dense: *ellipses* are frequent, *particles or emphatics* rare.

What of the 10 items that do not show much ability to predict technicality? These are: *increasing complexity in the content, grammatical writing, connotative language, examples, complex sentences, passive constructions*, and *statements of opinion or belief*. In addition, the three items which directly count instances all appear to be weak indicators: *terms defined in the text, non-textual elements* (such as figures), and *lists*. The inventory has been revised and these items removed from it.

Caution must be exercised in removing items, however. Consider #3, *citations*. Although this item passes the χ^2 test, its coefficient of correlation with technicality is 0.52, not a level usually considered significant. Yet most people would deem citations to be a very strong indicator of technicality. Why do the data not support this? A look at the sample texts and their overall technicality ratings provides an explanation. Of the five documents judged fairly technical, only one incorporates citations. This did not happen because the sample set was too small or ill-chosen. It is a fact that many of the texts we read every day—instructions and directions, information sheets—do *not* cite authorities but *are* fairly technical. It seems that there is more than one way for a text to be technical.

Examination of the 10 items does not provide any compelling reason to reconsider any deletion, however. Each was a plausible candidate to indicate technicality; it is also plausible for rating to have found this not to be so. Grammaticality, complex sentences, lists—it is possible to imagine both technical and non-technical texts which possess and lack these and the other features. *Terms defined in the text* is the one likely exception, and both it and the other two items which count instances require additional scrutiny.

8. Results

8.1. Revising the inventory

Table 10 presents the organization of the inventory revised in accordance with the results presented in the last section. The three columns on the right classify the remaining items. TYPE indicates the sort of rating involved (binary or scaled); SCOPE, the extent of text inspected as a basis for the rating (sentence, paragraph, page or entire document), and SPHERE, the linguistic category involved (lexical, syntactic, semantic or pragmatic).

The greatest change is the removal of the three items which count instances—all values are now either binary or scale ratings, yes/no or 1-in-5. The inventory continues to offer items which can be evaluated automatically, either easily or with some effort, as well as ones which only a person can adjudicate. This is fortuitous. A rating scheme entirely in the hands of the rater would be open to criticism as overly subjective even if it assessed a single ill-defined feature in terms of a number of more obvious ones. The retention of features whose instances can be counted—third person; present-tense verbs—in course of assessing their prevalence in the text provides an objective basis for any technicality rating based on the revised inventory. Contrariwise, an entirely automatic rating procedure is implausible; technicality is just too complicated a notion to be assessable without human intervention. All human beings are expert at using language. It is therefore both unsurprising and satisfying to see that the inventory retains many items that employ users' text assessment skills. A second satisfying outcome is that significant features were found in all linguistic spheres. Every dimension of a text contributes to the determination of its technicality.

Now to turn to a less positive result. The original rating procedure took 104 min to complete on average, a figure which includes the rating of a 16-*line* catalogue entry as well as a 19-page literary essay. This has been universally criticized as being too lengthy. There is unfortunately no obvious way to simplify matters. Sampling is already being used throughout the process. For instance, the 'essay' in question is actually a chapter from a 300-page book and therefore already a sample. Raters are asked to choose sample units whenever the number of units (pages, paragraphs, sentences) exceeds 10, and items have been moved down to consider the smallest unit consistent with getting a meaningful assessment. Nor can we look for a decrease in rating

Table 10. Revised technical text inventory

	Item title	Type	Scope	Sphere
1	an introduction or summary setting out the content?	binary	document	syntactic
2	a table of contents, index, appendices?	binary	document	syntactic
3	citations?	binary	document	lexical
4	quoted dialog?	binary	document	syntactic
5	special fonts and punctuation?	binary	document	lexical
6	title and headings follow a convention?	binary	document	pragmatic
7	does the text clearly identify its topic?	binary	document	pragmatic
8	is the text focused on its topic?	binary	document	pragmatic
9	does it communicate a body of knowledge?	binary	document	pragmatic
10	is the treatment serious and objective?	binary	document	pragmatic
11	is the presentation of ideas orderly and logical?	binary	document	pragmatic
12	is the material organized hierarchically?	binary	document	syntactic
13	are topics accessible at random?	binary	document	pragmatic
14	domain-specific terminology?	scalar	page	lexical
15	conventional expressions or formulae?	scalar	page	lexical
16	humor, sarcasm, invective?	scalar	page	pragmatic
17	colloquialisms?	scalar	page	lexical
18	vague terms?	scalar	paragraph	semantic
19	explicit analogs only?	scalar	paragraph	semantic
20	figurative language?	scalar	paragraph	semantic
21	ellipsis?	scalar	paragraph	syntactic
22	text binders, general hedges?	scalar	paragraph	syntactic
23	discourse particles, general emphatics?	scalar	paragraph	syntactic
24	use of the third person?	scalar	paragraph	syntactic
25	causative subordination?	scalar	paragraph	semantic
26	domain-specific verbs with single senses?	scalar	paragraph	semantic
27	verbs in present tense?	scalar	paragraph	syntactic
28	'be' sentences?	scalar	paragraph	syntactic
29	interrogative and imperative sentences?	scalar	paragraph	syntactic
30	generic reference?	scalar	sentence	semantic
31	unambiguous references?	scalar	sentence	semantic
32	nominalizations?	scalar	sentence	syntactic

time commensurate with the reduction in size of the revised inventory—the time spent sampling and counting syntactic constituents used in more than one item is unchanged. A saving of perhaps 10–15 min is more likely.

One action which would reduce the effort and time involved in rating significantly is to discontinue the practice of asking raters to count instances of an item in the sample text before assessing its prevalence there with respect to text in general. This request is intended to ensure that a rating is based on an accurate measure of the particular phenomenon's presence in the sample text. Dispensing with it might have a great impact on the accuracy of the rating. The question is easy to decide, however; rate texts both with and without counts and compare the results. Such an experiment would also indicate how much time would be saved by dropping counting. It cannot be suggested however that we believe that by dropping the discipline of making counts is likely to be satisfactory. Most likely accuracy will suffer unduly. A better solution to the problem must be found.

8.2. Using the inventory

The results achieved thus far accomplish one of the goals of this study—many researchers speak of using technical text without indicating what they mean, and we simply wanted to know if this rather ill-defined notion could be linked to more concrete features of a text. Even

without a conclusive analysis of the data we have learned that this is so. A related objective, to stimulate interest in and discussion of practical efforts to categorize texts, is advanced by submitting these results to public scrutiny. In that sense the study reported here can be viewed as a pilot for more comprehensive and exhaustive investigation of the manner in which these and other textual features express the character of the variety of texts of interest to researchers.

However we hope that going further at this time and providing a scheme to rate technicality in terms of the features which have proven to be meaningful predictors will also be useful, both to ourselves and to others who use texts they call technical. Researchers may want an objective rating for the various documents in a corpus, or they may simply want to order texts on this criterion. In so doing users may come to question our items or to identify likely features of a text which we have missed. Such reactions will affect the content and presentation of future versions of the inventory.

Therefore one final step remains—to construct a formula to compute a technicality rating from values of the 32 meaningful predictor variables. The most obvious approach is to solve a system of nine simultaneous linear equations in the 32 variables, taking as their coefficients the average ratings presented in Table 7; Table 11 recaps these averages for items shown to be indicators of technicality. The solution values would provide the weighting coefficients for a linear equation combining the inventory ratings into a single value.

Table 11. Normalized average values for meaningful inventory items, by text

1	introduction	1.0	2.3	1.0	2.3	4.0	2.1	2.6	2.6	4.7
2	ToC, Index	1.0	1.0	1.0	1.4	1.4	2.1	1.5	1.6	2.0
3	citations	1.2	1.2	1.0	2.8	1.4	1.0	1.2	1.4	5.0
4	quoted dialog	4.8	4.0	1.0	2.8	1.2	1.0	1.0	1.0	2.6
5	fonts & punctuation	4.3	2.4	1.6	3.5	3.5	4.5	3.4	5.0	5.0
6	titles, headings	2.1	1.3	1.6	1.7	3.9	3.3	3.2	3.7	5.0
7	topic identified	1.5	2.3	2.1	4.8	4.8	4.8	4.8	4.8	5.0
8	focus on topic	1.4	3.2	3.1	4.6	5.0	5.0	5.0	5.0	4.8
9	communicate knowledge	1.2	2.0	2.1	4.8	5.0	4.6	4.8	4.3	5.0
10	serious, objective	1.2	1.7	1.2	4.0	5.0	4.8	5.0	4.5	5.0
11	orderly & logical	1.2	2.9	1.8	3.7	5.0	4.3	4.6	4.2	5.0
12	hierarchic organization	1.0	1.2	1.0	1.5	4.6	2.1	3.2	3.4	4.8
13	random access	1.2	1.0	1.2	1.9	4.6	3.4	3.5	4.3	4.0
14	terminology	1.6	1.7	1.0	3.3	3.4	4.3	3.5	4.3	4.7
15	conventions	1.4	1.2	1.2	1.1	1.5	1.5	1.9	1.5	3.9
16	humor, invective	4.0	3.4	1.3	1.1	1.1	1.0	1.4	1.0	1.0
17	colloquialisms	4.2	3.2	1.5	1.4	1.4	1.1	1.4	1.0	1.0
18	vague terms	2.3	2.1	2.0	2.0	1.5	1.3	1.3	1.2	1.7
19	explicit analogs	1.3	2.0	2.2	2.0	1.3	1.0	1.2	1.3	1.6
20	figurative language	2.6	3.4	3.8	3.0	2.1	1.0	1.4	1.0	1.2
21	ellipsis	2.8	2.6	2.7	2.8	2.2	3.1	2.3	2.7	1.6
22	binders & hedges	1.7	2.5	1.9	2.4	2.7	1.2	1.6	1.5	3.4
23	particles, emphatics	2.5	2.6	1.6	2.2	1.8	1.0	1.2	1.0	1.9
24	third person	3.3	3.9	3.4	4.1	2.4	1.4	2.3	1.6	3.7
25	causative subordination	1.3	2.4	1.9	2.8	2.0	1.0	1.7	3.5	2.5
26	same-sense domain verbs	1.2	1.5	1.4	1.6	2.5	1.5	2.1	1.6	2.6
27	present tense	4.3	2.6	3.6	4.0	4.1	4.5	3.8	4.5	4.1
28	'be' sentences	2.9	2.8	1.9	3.3	2.9	1.2	1.3	1.8	2.9
29	interrogatives & imperatives	3.5	2.7	1.6	1.1	3.3	1.2	3.3	3.8	1.5
30	generic reference	1.9	2.2	2.9	2.4	3.6	2.8	2.6	3.4	3.2
31	unambiguous refs	1.8	2.5	2.2	2.2	2.5	2.4	2.4	2.8	2.0
32	nominalizations	1.3	1.9	2.1	1.7	2.5	2.4	1.2	2.3	2.5
	technicality	1.2	1.3	1.3	2.8	3.6	3.9	3.9	4.0	5.0

The number of equations in this set is less than the number of variables, so traditional methods of solving simultaneous linear equations, which require a square coefficient matrix, are not available to us. Instead linear programming must be employed. When using the simplex method this technique squares the coefficient matrix by shifting extra variables to the equation's right-hand side and setting their coefficients equal to zero, solving the reduced set, and then iterating over alternative combinations of variables to optimize the solution. Two consequences follow from this situation: (1) many solutions are possible (depending largely on which variables are in the matrix); and (2) a solution does not necessarily employ all variables.

The SOLVER[15] add-in supplied with Excel 5.0 employs linear programming techniques and is therefore suited to our problem. This equation solver repeatedly resolves[16] the set of rating equations using nine variables. Any formula directly based on a single such solution would ignore many of the items which appear to be meaningful indicators of technicality. This difficulty has been turned to an advantage by partitioning the set of items into three groups of nine and discarding the five items left over. Each of these subsets can be used to compute an independent technicality rating, or their solutions can be pooled on a one-third basis each to compose a rating based on 27 items. The groups have been chosen to emphasize a particular type of item. One contains items reflecting the text's *lexical* and *syntactic* features. A second includes *pragmatic* features that express the author's purpose in writing the text. The last group is more heterogeneous but holds a number of features that pertain to document *structure*.

The three groups follow. Items are listed in inventory order; the parenthesized number after an item is its position in the *revised* version of the inventory. The LEXICO-SYNTACTIC group is composed of:

1. *terminology* (#14),
2. *colloquialisms* (#17),
3. *binders & hedges* (#22),
4. *particles, emphatics* (#23),
5. *third person* (#24),
6. *present tense* (#27),
7. *'be' sentences* (#28),
8. *interrogatives & imperatives* (#29),
9. *nominalizations* (#32).

The PRAGMATIC group is made up of:

1. *topic identified* (#7),
2. *focus on topic* (#8),
3. *communicate knowledge* (#9),
4. *serious, objective* (#10),
5. *orderly & logical* (#11),
6. *hierarchic organization* (#12),
7. *random access* (#13),
8. *humor, invective* (#16),
9. *figurative language* (#20).

The STRUCTURAL group contains:

1. *ToC, index* (#2),
2. *citations* (#3),

3. *quoted dialog* (#4),
4. *fonts & punctuation* (#5),
5. *titles, headings* (#6),
6. *conventions* (#15),
7. *ellipsis* (#21),
8. *causative subordination* (#25),
9. *same-sense domain verbs* (#26).

These particular groupings have been chosen for their utility. Rating based solely on lexical and syntactic features holds out the prospect of being automatable; a program could provide a provisional assessment of technicality without human intervention. On the other hand a document's structure and pragmatic assumptions are the features most amenable to being assessed after a simple scanning of it. This grouping allows for the quickest rating (and addresses the problem identified at the end of the previous section).

An alternative partitioning would group together into a first tier the most highly correlated items or those accounting for the greatest amount of variance. A second tier would contain the next most meaningful nine items, etc. Pooling would be accomplished on the basis of the relative importance of each tier's items. This is an equally reasonable approach to take, but the practical considerations detailed above proved more compelling—we would like to see this inventory used so we have chosen the organization that is more suited to users.

Five items were dropped from the inventory list to reduce its number to an even multiple of nine. An alternative would have been to add the next four most meaningful items dropped from the original inventory and make up another group of nine. The considerations in our choice of which items to drop were: (a) their relative unimportance in statistical terms; and (b) the difficulty or effort involved in assessing them correctly. Items were therefore not deleted from the set of those which both correlate highly and which exceed the appropriate critical value, but instead from the least meaningful items identified by just one of the two analytic techniques. These were: *introduction* (#1), *vague terms* (#18), *explicit analogs* (#19), *generic reference* (#30), and *unambiguous references* (#31).

Simplex solution of the equation set provides coefficient weights for each of the three subsets of items. In the formulae that follow, variable subscripts identify the item they apply to in the inventory.

LEXICO-SYNTACTIC:

$$0.69\ X_{14} - 0.20\ X_{17} + 1.17\ X_{22} + 0.14\ X_{23} + 0.42\ X_{24} + 0.05\ X_{27} - 0.67\ X_{28} - 0.49\ X_{29} - 0.24\ X_{32} = TT_{LEXSYN}$$

PRAGMATIC:

$$-5.70\ X_7 + 3.17\ X_8 + 14.03\ X_9 - 10.17\ X_{10} - 0.74\ X_{11} - 0.01\ X_{12} + 1.34\ X_{13} + 3.09\ X_{16} - 4.76\ X_{20} = TT_{PRAG}$$

STRUCTURAL:

$$0.65\ X_2 - 0.23\ X_3 - 0.85\ X_4 + 1.47\ X_5 - 1.68\ X_6 + 1.03\ X_{15} - 0.19\ X_{21} - 0.16\ X_{25} + 1.19\ X_{26} = TT_{STRUCT}$$

An inspection of these three formulae shows that a technicality rating based on lexical and syntactic features is most strongly affected by *binders and hedges*. *Terminology* and *sentences involving 'be'* are next in importance. The greatest contributors to a pragmatic computation are *communicate knowledge* and *serious, objective* treatment of an *identified topic*. *Figurative language* and *humor, invective* rank below. The structural rating is based mostly on *fonts &*

punctuation, titles, heading and *same-sense domain verbs*, with *conventions* and *quoted dialog* coming next.

The relative scale of the coefficients gives an indication of the quality of the equation's fit to the data on which it is based. A variable outweighed substantially by one or more of its peers contributes less to the computation, and beyond some point its presence in the equation becomes questionable. According to this rule of thumb, the pragmatic solution is markedly inferior to the others. Its coefficients range over four orders of magnitude, 14.03 for *communicate knowledge* to 0.01 for *hierarchic organization*. Elsewhere the range is two orders (1.17–0.05) or one (1.68–0.19).

A technicality rating involving all 27 meaningful features on a equal basis would be:

$$\tfrac{1}{3}\,TT_{LEXSYN} + \tfrac{1}{3}\,TT_{PRAG} + \tfrac{1}{3}\,TT_{STRUCT} = TT$$

The multiplicity of formulae provided in this section serves to underline the provisional nature of the results reported here. The technicality formula is, at bottom, based on a partial analysis of some 200 ratings of nine texts. It is subject to change as additional analysis is performed on the existing data and further experimentation conducted. However, subject to these qualifications, we believe it can play a useful role in characterizing the commonly used but ill-defined concept of text technicality.

9. Future work

A number of tasks for the future follow directly from the work done to date. The most obvious is to perform the additional controlled administrations needed to test and revalidate the revised inventory. This might necessitate further adjustments to its contents and more testing. Next, new texts should be rated with the inventory to test the formula for computing a technicality rating. Multivariate statistical analysis must be completed and may have an impact on this line of activity. There are also a number of additional tasks that would strengthen the overall study.

- Data accumulated in the pre-study could be compared with that produced in the main study. Although the rating procedure changed somewhat between the two events, this change was evolutionary rather than revolutionary. If the pre-study data is consistent with the main body, incorporating it would add a tenth rather different document to the sample set, a chapter from an introductory accounting text.
- Although they have been labeled *'be' verbs*, etc., more accurate naming for a number of items in the inventory would reflect the fact that what is measured is the rater's impression of a given syntactic phenomenon. Subjectivity is introduced by two factors here: (1) the rater may or may not have an accurate impression of the feature's presence in the text; and (2) raters will certainly vary in their opinions about such a feature's occurrence in text in general.[17] The latter information is not readily measured. The former, however, can be checked against the counts of instances that raters were asked to make during the rating process. Although the duration of the process and the number of texts involved worked against the keeping of complete records, enough data is available to permit an investigation of how accurately raters perceived countable syntactic phenomena in the texts, and this is a suitable subject of a sub-study. Variance between raters' impressions of the frequency of a syntactic feature in a text should not exceed the variance in their opinions about its relative occurrence in the text, or the latter data risks being confounded.

- The rater population divides naturally into two groups, those who work in the KAML lab, and others who were recruited from a local Usenet conference. It would be interesting to compare the results achieved by these two distinct groups. Raters recorded the time they spent rating each text; also of interest would be comparisons between the results of those who took less than 40 min on average, those who took 40–100 min, and those who took more than 100 min. Did speed introduce any differences?

Douglas Biber's work is of great interest to us. His features overlap ours on the lexical and syntactic levels and offer the major advantage of being objective and therefore calculable in an automatic fashion (indeed, only the most highly motivated individual could face counting the words and characters needed to accumulate them). Although we do not believe that technicality can be predicted by features on these two textual levels alone, it would certainly be worthwhile to find out if we are wrong. It would also be interesting to apply the revised inventory to the Brown corpus, paralleling Cutting and Karlgren's application of a subset of Biber's features to the texts that compose it. This would produce technicality ratings for a corpus in common use.

Acknowledgements—The authors would like to thank Dr Belkacem Abdous in the Département de mathématiques et d'informatique at the Université du Québec à Trois-Rivières for his guidance on statistical analysis; Chris Drummond for reading and commenting on this paper; and the raters who provided the data on which the paper is based: Betty Jo Barker, Sabrina Barker, Matthew Barr, Kim Beaudry, Anna Bell, Margaret Bienert, Lynne Bowker, Paul Caron, Myrtle Daubney, Corina Dinescu, Ron Fritz, Earl Hagen, Stan Matwin, Dana McCarthy, Margaret Redding, Delphine Renie, Riverson Rios, Elizabeth Scarlett, Jean Stalker, Jon Torrance and Sylvester Zafiroff. These people spent the many hours documenting their intuitions and observations about texts which made this study possible.

NOTES

[1]'Scientific' is a near but not exact synonym. For simplicity, 'technical' and 'technicality' will be used exclusively in this paper.

[2]Some of the categories in the ICE classification appear to be genres while others are not.

[3]'Public' is not a synomyn for 'printed', nor 'private' for 'non-printed'. For our purposes however the equivalence holds. We are not interested in printed, private communications or in spoken language of any sort.

[4]Kittredge (1987) makes a similar argument, emphasizing the constraints on a sublanguage which render it computationally tractable.

[5]e.g. weather reports, horoscopes, recipes, product documentation, service manuals.

[6]e.g. letters, newspaper stories, technical abstracts, government regulations.

[7]By this term Hutchins means 'the numerous devices by which temporal, logical, causal and many other types of relationships between sentences or clauses are signalled', primarily conjunctions and sentence adverbs (like 'while', 'afterwards', 'because') that express temporal, spatial or causal connections. Hutchins remarks that 'there is no necessary correspondence between the sequence of sentences or clauses in a text and the logical or chronological sequence of their respective propositions, actions or events.'

[8]Narrative, descriptive, classificatory and evaluative.

[9]A classification of the roles of particular episodes in the development of the thought. A number of schemes have been proposed for fiction; Hutchins adapts them without difficulty to different categories of scientific papers.

[10]A sixth dimension, on-line informational elaboration, has not been used in subsequent analysis.

[11]See Biber (1989) for an evaluation and critique of previous linguistic typologies and a discussion of issues involved in the task.

[12]The page rather than the section was used because the former seemed to be a better classifier. Many documents are not well structured into sections, but there is a practical physical limit to the size a page can be.

[13]The objective was to assess the number of sentences in which 'be' was the main verb.

[14]Our experience of running an experiment over the Internet was positive. We would do it again. It provides an almost cost-free opportunity of tapping individuals who share a community of interest rather than geographic proximity (Usenet respondents had approximately the same post-graduate background as lab staff), and permitted quite close contact between experimenter and subject via eemail as necessary.

[15]SOLVER was developed for Microsoft by Frontline Systems, http://www.frontsys.com/. See their site for more information on its operation.

[16]Solutions were verified by substitution into the original equations; certain computations were duplicated using the LinPro v2.0 public domain linear programming tool.

[17]Human error is a third confounding factor.

REFERENCES

Arnold, D. (1989) Text typology and machine translation: an overview. *Translating and the Computer* **10**, 73–79.

Arnold, D., Balkan, L., Meijer, S., Humphreys, A. L., and Sadler, L. (1995) *Machine Translation: An Introductory Guide*. Online publication at http://clwww.essex.ac.uk/~doug/book/book.html.

Biber, D. (1988) *Variation Across Space and Writing*. University Press, Cambridge.

Biber, D. (1989) A typology of English texts. *Linguistics* **27**, 3–43.

Biber, D. (1993) Using register-diversified corpora for general language studies. *Computational Linguistics* **19**, 219–241.

Biber, D. (1995) *Dimensions of Register Variation: A cross-linguistic comparison*. Cambridge University Press, Cambridge.

Boggess, J. E. (1993) Using a neural network for syntactic classification of words in technical text. In Crocker, G. T. and Dean, S. T. (Eds), *Proceedings of the 31st Annual Southeast ACM Conference*. Association for Computing Machinery, New York.

Britton, B. K. and Graessner, A. (1975) *Text Comprehension and Understanding*. Erlbaum, Hillsdale, NJ.

Chervak, S. G., Drury, C. G. and Ouellette, J. P. (1996) Field evaluation of simplified English for aircraft workcards. In *Proceedings of the 10th FAA/AAM Meeting on Human Factors in Aviation Maintenance and Inspection*.

Cooper, C. R. and Matsuhashi, A. (1983) A theory of the writing process. In Martlew, M. (Ed.), *The Psychology of Written Language*, 3–39. Wiley, New York.

Cordova, J. L. (1992) *A Domain-Independent Approach to Knowledge Acquisition From Natural Language Texts*. Department of Computer Science, Mississippi State University.

Daily, J. E. (1971) Classification and categorization. In Kent, A. and Lancour, H. (Eds), *Subject and Information Analysis*, Readings from the *Encyclopedia of Library and Information Science*, 327–349. Dekker, New York.

Daily, J. E. (1976) Natural classification. In Kent, A. and Lancour, H. (Eds), *Subject and Information Analysis*, Readings from the *Encyclopedia of Library and Information Science*, pp. 350–369. Dekker, New York.

Dillon, A. and McKnight, C. (1990) Towards a classification of text types: a repertory grid approach. *International Journal of Man–Machine Studies* **33**, 623–636.

Frail, R. P. and Freedman, R. S. (1991) Text classification in fragmented sublanguage domains. Proceedings of the 7th IEEE Conference on Artificial Intelligence Applications, 1991, 33–36.

Gauch, S. and Futrelle, R. P. (1993) The Scientist's Assistant: structure-based navigation of technical texts. In *Proceedings of the Second Annual Symposium on Document Analysis and Information Retrieval*. University of Nevada at Las Vegas, 367–377.

Grishman, R. and Kittredge, R. (1986) *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Erlbaum, Hillsdale, NJ.

Hutchins, J. (1977) On the structure of scientific texts. UEA Papers in Linguistics, University of East Anglia.

Karlgren, J. (1995) *Identification of Text Genre and Context*. Online publication at http://www.sics.se./~jussi/distrnontopicir/distrnontopicir.html.

Karlgren, J. and Cutting, D. (1994) Recognizing text genres with simple metrics using discriminant analysis. In *Proceedings of the Fifteenth International Conference on Computational Linguistics (COLING 94)*.

Kilgarriff, A. (1995) *Review of dimensions of register variation: a cross-linguistic comparison. Journal of Natural Language Engineering* **1**, 396–399.

Kittredge, R. (1982) Variation and homogeneity of sublanguages. In Kittredge, R. and Lehrberger, J. (Eds.), *Sublanguage: Studies of Language in Restricted Semantic Domains*, 107–137. de Gruyter, Berlin.

Kittredge, R. (1987) The significance of sublanguage for automatic translation. In Nirenberg, S. (Ed.), *Machine Translation: Theoretical and Methodological Issues*, 59–67. Cambridge University Press, Cambridge.

Kniffen, J. D., Kincaid, J. P., Lang, S. and Thomas, M. (1989) Computer aids for authoring technical text written in controlled English. *Proceedings of the Human Factors Society* **33**, 1134–1138.

Kontos, J. and Cavouras, J. C. (1988) Knowledge acquisition from technical texts using attribute grammars. *The Computer Journal* **31**, 525–530.

Lai, P. and Manber, U. (1991) Flying through hypertext. In *Proceedings of the ACM Hypertext '91 Conference*.

Lehrberger, J. (1986) Sublanguage analysis. In Grishman, R. and Kittredge, R. (Eds.), *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*, 18–35. Erlbaum, Hillsdale, NJ.

N.Zafio, M. (1995) Ontologie des textes dans un discours de spécialité: le pari de l'immanence textuelle, Ph.D. thesis, Département de linguistique et de traduction, Université de Montréal.

Obermeier, K. K. (1985) GROK—a knowledge-based text processing system. *Proceedings of the Artificial Intelligence and Advanced Computer Technology Conference*, Weisbaden, 384–389.

Paice, C. D. and Jones, P. A. (1993) The identification of important concepts in highly structured technical papers. In *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 69–78.

Roudaud, R. (1992) Typology study of French technical texts with a view to developing a machine translation system. In *Proceedings of the Thirteenth International Conference on Computational Linguistics (COLING 92)*, 1284–1288.

Rowe, N. C. and Laitinen, K. (1995) Semiautomatic disabbreviation of technical text. *Information Processing & Management* **31**, 851–857.

Sager, J. C. (1982) Types of translation and text forms in the environment of machine translation. In Lawson, V. (Ed.), *Practical Experience of Machine Translations*, 11–19. North-Holland.

Survey of English Usage. (1991) International Corpus of English (ICE). University College, London.

Sutcliffe, R. F. E., Brehony, T. and McElligott, A. (1994) Link grammars and structural ambiguity: a study of technical text. Department of Computer Science, University of Limerick. Technical Report UL-CSIS-94-16.

Sutcliffe, R. F. E., Brehony, T. and McElligott, A. (1994) The grammatical analysis of technical texts using a link parser. Department of Computer Science, University of Limerick. Technical Report UL-CSIS-94-14.

Vickery, B. C. (1968) Analysis of information. In Kent, A. and Lancour, H. (Eds), *Subject and Information Analysis*, Readings from the *Encyclopedia of Library and Information Science*, 6–35. Dekker, New York.

Wenger, M. J. (1993) Reduced text structure at 2 text levels—impacts on the performance of technical readers. *Journal of Technical Writing and Communication* **23**, 333–352.

KEN BARKER is a doctoral student in computational linguistics at the University of Ottawa. His thesis deals with the semantic analysis of English technical text in the absence of precoded semantic knowledge. He has published two journal and three conference papers.

TERRY COPECK is a research associate at the KAML laboratory of the University of Ottawa Department of Computer Science with interests in semantic analysis and text processing.

SYLVAIN DELISLE is a professor of Computer Science at l'Université du Quebec à Trois-Rivières. His current research interests include syntactic and semantic analysis of text, computational linguistics and artificial intelligence.

JEAN-FRANCOIS DELANNOY received a Ph.D. in computer science from the University of Marseilles and is now a knowledge engineer at RES International, Ottawa, working on information retrieval and on text summarization. He is interested in technical, cognitive and philosophical aspects of understanding and communication.

STAN SZPAKOWICZ is a professor of Computer Science at the University of Ottawa, Canada. His current research interests include text analysis for knowledge acquisition. He has published over 80 refereed papers and six books.

APPENDIX

Text Technicality

/
HR MIN

_____ / _____
TITLE RATER

Choose 10 pages from documents longer than ten pages. Pick 10 paragraphs from these pages or directly from shorter documents. Pick 10 sentences from the paragraphs or directly from a text less than ten paragraphs long. Record your choices below. See the manual for more information on sampling.

Sample												
NOT SAMPLED ☐	1	___	___	___	2	___	___	___	3	___	___	___
	4	___	___	___	5	___	___	___	6	___	___	___
7	___	___	___	8	___	___	___	9	___	___	___	10	___	___	___
	PAGE	PARA	SENT	PAGE	PARA	SENT	PAGE	PARA	SENT	PAGE	PARA	SENT

Take any sample pages out of the second copy of the document. Draw a bar besides the selected paragraph on these pages and highlight the chosen sentence within it.

Entire Document

Read the document. Answer the following questions based on your observations of and conclusions about the *entire* text. Turn to the manual for more information on how to assess particular items.

YES/NO YES NO

COMPONENTS

1. **an introduction or a summary setting out the content?** ___ / ___

   Does a section or subsection recap the document's message? This can be an Introduction, Summary or Conclusion, or items particular to the subject at hand such as a list of ingredients in a cookbook.

2. **a table of contents, index, appendices?** ___ / ___

   Is the document augmented with one or more of: table of contents; list of tables, figures, illustrations; glossary, index, appendices? Do not count a colophon, dedication, frontispiece or postscript.

3. **citations?** ___ / ___

   Citations in the body like [Smith, 1993] or a list of references at the end qualify. A list of recommended readings would be counted under the previous item.

4. **quoted dialog?** ___ / ___

   Any repeated use counts. Direct or verbatim quotations are usually set off with single or double quotes and often accompanied by a reference to the speaker: "Don't be silly!", she said.

5. **special fonts and punctuation?** ___ / ___

   A catch-all for everything from mathematical formulae and card suits in bridge hands, to the use of a different typeface to set off 'Don't be silly' in item 4. Instances are usually easy to detect.

6. **title and headings follow a convention?** ___ / ___

   Look for Introductory, Proceeding, Collected Papers, Manual, Workbook etc in titles; for Results, Methodology or prefixed runs of letters, roman or arabic numerals in chapter and section headings.

CONTENT

7. **does the text clearly identify its topic?** ___ / ___

   Does the text identify the *topic* outright if it isn't obvious? Is the *purpose* of the document mentioned explicitly—a textbook is introductory; a paper surveys a topic; a manual deals with an upgrade. Does the author indicate the *scope* and *limit* of his treatment, perhaps by contrasting it with other texts?

YES NO

8. **is the text focused on its topic?** ___ / ___

Does discussion stay on topic or introduce material extraneous to the author's purpose? Conversely, does it give enough information to cover the topic convincingly? Are any obvious issues skipped?

9. **does it communicate a body of knowledge?** ___ / ___

Does the text add something to the discussion of its topic, either factual information or an interpretation of facts? Can you summarize its message other than to say 'this is a book about..'?

10. **is the treatment serious and objective?** ___ / ___

Serious treatment means trying to communicate a message rather than to entertain or to amuse. *Objective treatment* means the author attempts to present facts independent of his beliefs or opinions. Objectivity does not mean that a text cannot contain opinion, just that it distinguishes statements of belief from statements of fact.

ORGANIZATION

11. **is the presentation of ideas orderly and logical?** ___ / ___

Does the discussion begin with definitions or fundamental concepts and build from there? Is the presence or location of any idea surprising? Are arguments complete and coherent? Does the development suggest thought and planning?

12. **does discussion become more complex as it proceeds?** ___ / ___

Gauge this by checking whether concepts introduced towards the end specialize more general ones presented earlier in the text. Consider several; write down those you check in the margin of this form.

13. **is the material organized hierarchically?** ___ / ___

Check the organization of sections within chapters, of subsections within sections. Do the subjects of subordinate units fall within the scope of their parent?

14. **are topics accessible at random?** ___ / ___

After reading the text can you turn directly to the approximate place where a given concept is discussed?

15. **is writing grammatical, are sentences well-formed?** ___ / ___

Do you have to re-read sentences to get their meaning? Does the discussion flow? Is the author's style obtrusive?

Page Samples

Set the original document aside and turn to the sample pages, if any. The rest of the rating uses samples exclusively: first pages, then paragraphs and finally sentences. Look at the *pages* now and count how often features mentioned in the next three items occur, recording examples on the adjacent dotted line as requested. Enter the total on the line at the end of the item.

Counts

16. **terms defined in the text:** Look for terms defined by the author. Count occurrences; the same term appearing twice would count as two. Defined terms often appear in an index and their first occurrence may be styled to stand out from the body of the text. Is there a glossary?

i.e. ..... _____

17. **non-textual elements:** Non-textual items are drawings, photographs, figures, tables etc—anything that supplements the text. Note the various kinds on the dotted line. Do not count lists.

kinds ..... _____

18. **lists:** Lists may be embedded in the body or broken out into paragraphs. They often start with an introducer that ends in a colon; individual points may begin with a bullet, long dash or incrementing key. Points can be entire sentences, clauses or phrases, and can end with commas or semicolons, or without punctuation. The chosen style is used consistently throughout the list.

_____

Scales

To grade an aspect of the text on a scale, compare it against your impression of that feature's prevalence in writing in general and check the scale. If it is

— much more common than usual, check thus: PRESENT | √ | | | | | ABSENT

— about as common as usual, check thus: PRESENT | | | √ | | | ABSENT

— *much less* common than usual, check thus: PRESENT | | | | | √ | ABSENT

The other two grades are *somewhat more* or *somewhat less* common than usual. We ask you to count instances in some items. These are meant to give a basis for whatever grade you pick. Note observations or examples that add substance to your rating on the lines below an item. Copy short examples entirely; identify long ones in page/paragraph/sentence form. Rate any item that is inappropriate to the text 'absent'.

19. domain-specific terminology?

PRESENT | | | | | | ABSENT

Count all domain-specific terms in the sample. These can be common or proper names, abbreviations like NASA or even symbols: TEX. They may be inventions: firmware or existing words reused in a sense peculiar to the domain: function in mathematics. Most terms are nouns or noun phrases, but be alert for the occasional verb or modifier. Include *all* domain-specific terms, not just those defined in the text. Record instances and their total number and rate the text.

total domain-specific ..... instances .....

20. conventional expressions or formulae?

PRESENT | | | | | | ABSENT

A formal proof often begins with consider and assume and ends with a box: □. Academic writing commonly uses italicized Latin abbreviations in connection with citations: ibid, op. cit., et al. Describe in a few words any convention you encounter and rate accordingly.

total conventions ..... instances of convention .....

.....

21. humor, sarcasm, invective?

PRESENT | | | | | | ABSENT

These literary devices are not marked in any consistent way. Record page/para/sent details and a word or phrase identifying each instance. Count them and evaluate the total to rate the text.

total ..... instances .....

.....

Paragraph Samples

Take a pencil and underline all verbs in your *paragraph* samples with straight (or colored) lines. Count them; this count will be used in a number of ratings to come.

total verbs .....

22. colloquialisms?

PRESENT | | | | | | ABSENT

Count colloquial, idiomatic and informal phrases like go figure and contractions like isn't. Record the counts and the instances of colloquialisms. Rate the text in the light of these totals.

total contractions ..... total colloquialisms ..... instances of colloquialisms .....

.....

23. **vague terms?** PRESENT |_|_|_|_|_|_| ABSENT

Count instances of terms like thing, idea, action etc. Ignore such terms when they occur in idioms. Record instances and the total and rate the text.

total vague terms ..... instances of vague terms .....

.....

24. **meaning conveyed by denotation?** PRESENT |_|_|_|_|_|_| ABSENT

Investigate this negatively—look for words or phrases with connotations. A sentence beginning with obviously may beg the question; compare file actor to movie star Record each instance and the total number, then rate the text.

total connotative words/phrases ..... instances of connotation .....

.....

25. **explicit analogs only?** PRESENT |_|_|_|_|_|_| ABSENT

Decide whether candidates figures are analogies or similes and whether the analogous relationship is clearly identified. Only if both conditions are true should this feature be rated present. Similies are comparisons involving two unlike thing: my love is like a red, red rose. Analogies are extended similes spanning a number of sentences; analogs usually resemble each other on more than one aspect or in a deep rather than a surface way. Both may be introduced by like or as. Record the total and a word identifying each explicit analogy. Rate the text.

total analogies ..... total explicit ..... instances of explicit analogy .....

.....

26. **figurative language?** PRESENT |_|_|_|_|_|_| ABSENT

Similes may be introduced by like or as while metaphors are just used in place: a glaring error. Each is a single phrase. See the manual for other less common figures. Count the number of figures; circle them in the text to record them. Enter the total below and rate the text.

total figures .....

27. **examples?** PRESENT |_|_|_|_|_|_| ABSENT

Examples are often preceded by e.g. or set out from the body of the text in some way. Record each in page/para/sent notation along with a key word to identify it. Rate their frequency in the text.

total examples ..... instances of examples .....

.....

28. **ellipsis?** PRESENT |_|_|_|_|_|_| ABSENT

Ellipsis tends to occur in repeated structures; nouns, verbs or prepositions are dropped from later elements in the series. Record instances by adding a 'V' insert mark where words have been dropped and rate the text according to your overall impression.

total ellipses .....

29. **text binders, general hedges?** PRESENT |_|_|_|_|_|_| ABSENT

Count and record instances of sequencers like first, second, .. and link words like next, however etc. Binders and hedges often appear at the beginning or end of sentences, or between clauses. Rate the text. in the light of your findings.

total ..... binder & hedge instances .....

.....

30. **discourse particles, general emphatics?** PRESENT |_____| ABSENT

Count discourse particles (isn't it?, eh, as a matter of fact) and emphasizers (certainly, clearly, without a doubt). Evaluate these counts to rate the text.

total discourse particles ..... total emphatics .....

31. **use of the third person?** PRESENT |_____| ABSENT

Record the total number of pronouns and the number in third person. Ignore the singular/plural distinction. Rate the text in the light of these counts.

total pronouns ..... total 3rd person .....

32. **causative subordination?** PRESENT |_____| ABSENT

Count the number of subordinate or coordinate sentences that use words marking a cause-effect relationship: since, because, if...then, as (conjunction), when etc. Enter the total and grade the text.

total sentences ..... total causal .....

33. **complex sentence structure?** PRESENT |_____| ABSENT

Classify sentences as either simple or complex. Simple sentences have a single clause, complex ones have two or more. Grade the text in the light of your counts.

total simple ..... total complex .....

34. **domain-specific verbs with single senses?** PRESENT |_____| ABSENT

Record verbs that are used in a sense particular to the domain of discourse. Ascertain if each has a second more general sense. Use a dictionary as necessary. Record the number of domain-specific and multiple-sense domain-specific verbs and instances of the latter and rate the text accordingly.

total domain-specific ..... total multiple-sense domain-specific .....

instances of multiple-sense .....

35. **verbs in present tense?** PRESENT |_____| ABSENT

Count the number of verbs in present tense. Ignore modals and the perfect/progressive distinction. Compare this figure to the total number of verbs and rate the text.

total verbs ..... total present tense .....

36. **'be' sentences?** PRESENT |_____| ABSENT

Count the number of verbs which are some form of be or become. Ignore other statives like seem, feel etc. Evaluate this figure and the total number of verbs and rate the text.

total verbs ..... total stative verbs .....

37. **passive constructions?** PRESENT |_____| ABSENT

Count the number of verbs in passive voice: I was seen vs I see. Compare this figure to the total number of verbs and rate the text.

total verbs ..... total passive verbs .....

38. **interrogative and imperative sentences?** PRESENT |_____| ABSENT

Questions end with a ?. Commands may end with an ! and often begin with a verb; they frequently appear in a series. Record your count of each kind of sentence and evaluate the text.

total questions ..... total commands .....