Discourse Constraint in Computer Manuals

Tetsuya Nasukawa

IBM Research, Tokyo Research Laboratory
1623-14, Shimotsuruma, Yamato-shi, Kanagawa-ken 242, Japan
Phone: 81-462-73-4574, 81-462-73-7413(FAX)
nasukawa@trl.vnet.ibm.com

Abstract

This paper describes the result of our analysis of computer manuals to evaluate the strength of discourse constraint, due to which morphologically identical words tend to be used in the same way in the same context. According to our analysis, more than 90% of all the content words in a single chapter of a computer manual have the same word sense. For a manual concerning a narrow domain, such as a manual for experts, this constraint is stronger, but it becomes weaker with in a larger scope such as that of the whole manual, where 92.6% of the verbs have the same word sense, as opposed to 96.3% within a single chapter. This discourse constraint can be used to solve various problems in natural language processing, such as lexical and structural disambiguation and pronoun resolution. The results of this analysis offer a good prospect for the possibility of developing practical method of context processing that improves the accuracy of sentence analysis. However, generation of a good translation in a target language also involves the issue of free translation. In our analysis, free translation was detected in 35.5% of verbs in a manual professionally translated into Japanese.

1 Introduction

Improving the accuracy with which lexical, attachment, and various other ambiguities can be resolved in sentence analysis is one of the main issues in natural language processing. Various approaches have been proposed, including the case-based approach [8] and the statistically based approach [2]. Most of these previous approaches relied on knowledge resources, and were limited by a knowledge acquisition bottleneck. Thus, the focus of recent approaches has been shifted to an attempt to relieve the knowledge acquisition bottleneck by trying to reduce human intervention in the construction of knowledge resources [3]. Some other approaches construct and refer to the context structure in order to improve the accuracy of text analysis; however, they have been considered high-cost procedures that require enormous amounts of background knowledge and deep inference mechanisms.

We have been developing a practical method of context processing by focusing on extracting inter-sentential information from the syntactic relations of morphologically identical words in different sentences in the same text. In a text concerning a restricted domain, such as a computer manual, the vocabulary is relatively small, and morphologically identical words are frequently repeated. By hypothesizing that morphologically identical words are used in the same way (namely, that they are used in the same word sense, and modify or are modified by similar concepts), we can apply this hypothesis as a discourse constraint (word sense consistency) and discourse preference (modification similarity) in order to solve various kinds of ambiguity in sentence analysis. This approach is practical in terms of the amount of knowledge it presupposes and the amount of computation it requires, since it basically relies only on the surface information in a text, and is free from the knowledge acquisition bottleneck.

In this paper, we describe the result of our analysis of English computer manuals to show that there is actually a strong tendency for morphologically identical words to be used in the same way. and investigate the size for the scope of a text that allows the most efficient application of the discourse constraint. As well as analyzing the discourse constraint, we examined how each word in the English manuals was translated into a Japanese expression by a professional translator, and found that out of all the lemmas that appear more than once in the same chapter with the same word sense, 7.1% of nouns and 35.5% of verbs were translated into various Japanese expressions in such a way that they can be considered to be the results of free translation. This indicates that much work is required on the generation phase of MT in order to achieve sufficiently high quality for publication; it also indicates that there are some limitations in the approaches that uses bilingual corpora as knowledge resources [6] [9].

In the next section, we describe how we analyzed computer manuals to evaluate the strength of the discourse constraint, and investigate the most appropriate size of the scope for application of the discourse constraint to the disambiguation process described in the following section. Section 3 describes how we can apply the discourse constraint and discourse preference to various problems in natural language processing, such as word sense disambiguation, attachment disambiguation, and pronoun resolution. In Section 4, we describe the results of our analysis of professional translations, examining how each word was translated into Japanese by a professional translator.

2 Analysis of Computer Manuals

2.1 Text Database

We have been building a text database containing two computer manuals. One [10] is a typical computer manual for computer experts such as programmers and system operators, and the other [11] is a primer for new users of a computer.

Figure 1 shows the structure of this database. It consists of original English sentences, their parsed tree output by the PEG parser [5], and their corresponding Japanese sentences professionally translated for publication in Japan. Question marks in the parsed tree indicate attachment ambiguities. For example, the question mark at the left of the prepositional phrase, "(PP (PP (PREP* "in"))" indicates that *in Journal Files* can be a modifier of the verb *Storing* "(VERB*"storing" ("store" PG))," which is located in the same vertical level as the question mark, as well as a modifier of *Messages* "(NOUN* "Messages" ("message" PL))," which is located in the same vertical level as the prepositional phrase "(PP (PP (PREP* "in"))." Thus, the parsed trees in this database hold attachment ambiguities and lexical ambiguities (since there is no information on word sense).

and this database functions as a workbench for various kinds of experiment on disambiguation.

Using this database, we extracted sentences for each word in each part of speech, and analyzed how morphologically identical words were used in each chapter of the same manual.

2.2 Analysis of Discourse Constraint

The precise data obtained in our analysis of verbs and nouns in each chapter of the typical computer manual are given in Table 1.[1] Numerals in parentheses in "Repeated Words" show the number of polysemous words [2] that appear more than once in the same chapter, and the strength of the discourse constraint is calculated according to the following formula:

$$\frac{\textit{(number of repeated polysemous words)} - \textit{(number of multiple uses of word sense)}}{\textit{(number of repeated polysemous words)}}$$

As in the table, the discourse constraint for nouns in the typical computer manual is quite strong, and only a small number of words break this constraint. All are basic words such as *guide, term, time, number, service,* and *source.*

Figure 2 shows data on the use of multiple word sense for *number* in Chapter 4, which we extracted from the text database. Here the word *number* is used with three different meanings:

1. A position in a series (Sentence 88)
2. An amount (Sentence 143, Sentence 144, Sentence 147)
3. A numeral (Sentence 209)

Figure 2 gives us the impression that writers may not intend to unify the senses of the above basic words, since these senses may be relatively easily distinguished by checking syntactic information. However, the words still tend to be used in a particular sense according to the topic or the context. For example, the word *number* is used seven times in Chapter 2, each time in the sense of "an amount." In Chapter 5, *number* also appears seven times, each time in the sense of "a position in a series." Thus, there is a tendency for every word to have a unified sense in the same discourse.

Table 2 shows the discourse constraint in a manual for beginners. This manual gives general information on computers; therefore, the domain of the topics in this manual is much broader than that in a typical computer manual for experts. Comparison with Table 1 shows that the discourse constraint is stronger in the manual for experts, where 97.8% of nouns and 96.3% of verbs in the same chapter take the same word sense, than in the manual for beginners, where 91.8% of nouns and 94.7% of verbs in the same chapter take the same word sense. This result indicates that the discourse constraint is stronger in a text concerning a narrower domain.

Table 3 shows the strength of the discourse constraint on verbs in the typical computer manual for experts within the scope of the whole manual. Comparison with the data in Table 1 indicates that the discourse constraint within the scope of a whole manual (92.6% word sense consistency)

[1] Numerals in parentheses in "Number of Multiple Uses of Word Sense" in "Verb" data indicate the numbers of words that are used with multiple senses in the same chapter but only as a part of idioms, and that can be easily distinguished from others by checking their syntactic information.

"Polysemy of each word is based on the number of definitions given in LDOCE [12].

Table 1: Discourse Constraint on Nouns and Verbs in the Manual for Experts

	Text Size		Noun				
	Num. of Sents.	Num. of Words	Num. of Words	Repeated Words		No. of Mlt. Uses of W. Sense	Strength of Discourse Constraint
				Number	Ratio		
Notices	57	692	273	48 (22)	68.5%	0	100%
AboutThisBook	196	1693	645	83 (45)	77.2%	2	95.6%
Chapter 1	123	1315	552	71 (27)	74.8%	1	96.3%
Chapter 2	256	3499	1336	173 (83)	84.1%	0	100%
Chapter 3	438	2287	1124	110 (42)	83.2%	2	95.2%
Chapter 4	277	2676	1075	106 (43)	90.2%	1	97.7%
Chapter 5	41	502	185	40 (21)	76.8%	0	100%
Chapter 6	93	964	353	47 (22)	80.5%	0	100%
Chapter 7	439	1998	937	89 (31)	87.2%	1	96.8%
Chapter 8	296	1829	774	105 (41)	69.3%	1	97.6%
Chapter 9	373	2908	1211	154 (78)	86.5%	3	96.2%
Total	2589	20363	Average				97.8%

	Verb				
	Number of Words	Repeated Words		Num. of Mult. Uses of Word Sense	Strength of Disc. Const.
		Number	Ratio		
Notices	65	15 (14)	64.6%	1	92.9%
About This Book	135	24 (22)	74.1%	3 (1)	86.4%
Chapter 1	116	26 (23)	69.8%	0	100%
Chapter 2	359	54 (46)	86.6%	4	91.3%
Chapter 3	153	30 (25)	85.0%	0	100%
Chapter 4	274	40 (34)	89.4%	1 (2)	97.1%
Chapter 5	44	9 (7)	63.6%	0	100%
Chapter 6	90	18 (15)	65.6%	1	93.3%
Chapter 7	138	22 (19)	82.6%	0	100%
Chapter 8	161	28 (27)	65.2%	0 (1)	100%
Chapter 9	302	55 (46)	82.1%	1	97.8%
	Average				96.3%

Table 2: Discourse Constraint in the Manual for Beginners

Part of Speech	Ratio of Repeated Words	Strength of Discourse Constraint
Noun	77.0%	91.8%
Verb	69.5%	94.7%
Adjective	73.3%	100%
Adverb	46.0%	100%

Table 3: Discourse Constraint on Verbs within the Scope of the Whole Text of the Manual for Experts

Number of Words	Repeated Words		Multiple Use of Word Sense	Strength of Discourse Constraint
	Number	Ratio		
1837	182 (149)	94.9%	11	92.6 %

is much weaker than the discourse constraint within the scope of a chapter (96.3% word sense consistency).

In order to investigate the size for the scope of a text that allows the discourse constraint to be most efficiently applied, we should consider both the strength of the discourse constraint, which is stronger in a smaller text, and the frequency of morphologically identical words, which is higher in a larger text, so that more information for disambiguation can be extracted from the context. As shown in Table 1, chapters with fewer than 2000 words have relatively low frequencies of morphologically identical words (under 80%), and in terms of efficient application of the discourse constraint for a disambiguation process that we describe in the next section, a higher frequency of morphologically identical words ensures a higher accuracy of disambiguation; thus, we presume that a chapter containing around 2000 words is the most appropriate scope for a discourse.

3 Context Processing Based on the Discourse Constraint and the Discourse Preference

In this section, we describe how the discourse constraint can be used to solve various problems in natural language processing.

3.1 Word Sense Disambiguation

Most previous research on word sense disambiguation has been focused on semantic consistency, which is calculated by using lexical resources such as an example-base [8], and usually applied within a sentence. This work has been limited by at least two factors: (1) a knowledge acquisition bottleneck and (2) a lack of contextual information.
For example, consider the phrase

(3.1) *Role of an operator.*

This kind of phrase is often used as a section heading. It does not contain sufficient information for disambiguation of the word sense of *operator.* The word *operator* can be (1) a person who operates a machine, or (2) a logical operator for computation, since the word *role is* semantically consistent for both senses of *operator.* In this case, the word *operator* is likely to be used in other sentences in the same text, and some of them might contain sufficient information to allow disambiguation of its word sense; for example,

(3.2) *Contact the operator to find out...*

(3.3) *The operator is responsible for the daily operations of the...*

(3.4) *Operators should keep this guide handy to use as a reference...*

When the discourse constraint is applied so that all the morphologically identical words have the same sense, the sense of *operator* in (3.1) reflects the result of disambiguation of other sentences that may contain sufficient information to allow disambiguation of the word sense of *operator.*

This approach is not a solution for the knowledge acquisition bottleneck. It requires a previous method in order to disambiguate word senses in at least one of the sentences, but it improves the overall possibility of obtaining a correct word sense. In the conventional approach, the possibility of correctly disambiguating the word sense of *operator* in sentences (3.2) - (3.4) above depends on the lexical resource. In our approach, as long as the lexical resource contains sufficient information for disambiguation of *operator* in one of the sentences, the result of this disambiguation can be referred to by all the other sentences. In this respect, our approach relieves the knowledge acquisition bottleneck. Moreover, for practical applications in natural language processing, a lack of consistency in the interpretation of morphologically identical words often confuses users.

3.2 Structural Disambiguation

The previous methods of structural disambiguation have much the same limitations as word sense disambiguation. They also depend on lexical resources such as example-bases [8] or on-line dictionaries [1], and can usually be applied only within a sentence.

In order to identify the importance of context for attachment ambiguities, consider the following well-known example:

(3.5) *He saw the girl with a telescope.*

In this sentence, the modifiee of *with a telescope* can be *saw* or *the girl,* depending on the context. If a sentence such as

(3.6) *He bought the telescope last weekend.*

follows (3.5), the information on *the telescope* in this sentence provides information that helps to determine the modifiee of *with a telescope.*

The next example is extracted from the same page of a computer manual: [3]

[3] Some modifiers have been omitted to simplify these sentences for illustrative purposes.

(3.7) *Printing documents in batch*

(3.8) *Printing in batch means to print a large number of documents in one stack.*

(3.9) *The documents are collected in the libraries until it is convenient for you to print them.*

Sentence (3.7) is actually the title of a chapter. In this phrase, *in batch* has an attachment ambiguity, since it may modify either *Printing* or *documents.* In this case, the information in (3.8) can be referred to, since it also contains *batch.* Because the modifiee of *in batch* in (3.8) is uniquely determined as *printing,* application of the discourse preference so that morphologically identical words in the same discourse modify or are modified by similar concepts determines *printing* to be the modifiee of *in batch* in (3.7).

3.3 Pronoun Resolution

In this paper, we divide pronoun resolution into two steps:

1. Selection of candidates for an antecedent, using syntactic information
2. Selection of the correct antecedent from the candidates, using semantic information.

Here, we do not consider cases in which the antecedent does not explicitly appear in a previous sentence. In fact, we could not find such a case in the computer manuals that we analyzed.

Step 1 can be performed with syntactic information by checking the number and gender of each word in previous sentences, and the order of preference among the candidates can also be determined from syntactic information, as proposed by Hobbs [4] and Sidner [7]. In most cases, the number of candidates selected in this step is relatively small as long as candidates are searched for among previous words in the same sentence or the preceding sentence; therefore, a simple filter for checking number and gender may be sufficient for this step.

Step 2 requires semantic information, since each candidate selected in Step 1 can be the antecedent, at least as regards syntactic consistency. Here again, previous methods for determining semantic consistency have much the same limitations as those we described for word sense disambiguation. Our proposal is to apply the discourse constraint and the discourse preference in Step 2 in order to select the most probable antecedent from among the candidates.

Consider, for instance, sentence (3.9) in the previous section:

The documents are collected in the libraries until it is convenient for you to print them.

The antecedent of *them* can be either *documents* or *libraries,* as both are selected in Step 1. according to the syntactic information. In order to determine which candidate is the correct antecedent, we should know which is appropriate as an object of *print.* To select *documents* in a conventional approach, a knowledge resource must contain some description indicating that *documents* can be printed. For example, if semantic features are used, the semantic feature of *documents* should be contained in a possible object list of *print,* and this list should not contain the semantic feature of *libraries.* In another case, there might be some example that indicates a preferable connection between *print* and *documents* rather than *print* and *libraries.*

In our approach, we collect all the sentences and phrases that contain *documents* in the same text. In this case, both sentences (3.7) and (3.8) are collected. Phrase (3.7),

Printing documents in batch,

contains explicit information that *documents* can be an object of *print* in both of the phrases: thus. through application of the discourse preference, *documents* is selected as the most likely antecedent of *print* in (3.9).

4 Analysis of Human Professional Translation

In this section, we describe the results of our analysis of professional translations. By using the same data as for the analysis of the discourse constraint, we examined how each word was translated into Japanese.

Table 4 shows the numbers of difference in the translations of morphologically identical words that appear more than once in the same chapter with the same sense. From this result, it is clear that morphologically identical words used in same sense may not be always translated into an identical expression in the target language; in particular, predicates tend to be varied when they are translated into a target language. To be precise, 7.1% of nouns and 35.5% of verbs of morphologically identical words repeated in the same chapter with the same senses were not translated into the same expressions in Japanese; moreover, they were often translated into entirely different expressions in Japanese that cannot be regarded as having counterparts in the source sentences without considering the discourse or background knowledge.

One of the two computer manuals we analyzed, the typical computer manual for experts, was translated by using an output of machine translation; thus, the translator began by editing a text in which morphologically identical English words used in the same sense were translated into identical Japanese expressions; therefore, the degree of variation in expression indicates the amount of rewriting done by the human translator. This indicates that no matter how accurately an MT system can analyze a source text, the generation mechanisms of current MT systems are limited in their capability to generate texts of sufficiently high quality for publication.

One reason why such free translation is required is the difference in the granularity of concepts covered by expressions in each language. For example, a single concept of the English word *wear*, "to have (esp. clothes) on the body," has to be interpreted by using one of various expressions in Japanese according to the part of the body on which the article is worn. The expression can be "被る (*kaburu*)" for the head, "履く (*haku*)" for the foot, and so on. Another reason, which we consider more essential, is the tendency of humans to change expressions for predicates.

During our analysis of English computer manuals, in which our aim was to determine the scope of *also,* which basically consists of comparison of the correspondence of concepts in two sentences, we found that most nouns were repeated with the same expressions unless they were replaced with pronouns or definite expressions such as *this, that,* and *the;* in comparison, predicates were often repeated with different expressions. For example:

A *has* B. → A *also includes C.*

A *contains* B. → C *is also included in* A.

Table 4: Analysis of Professional Translation of Nouns and Verbs in the Manual for Experts

	Noun					
	Number of Words					Ratio of Free Translation
	Repeated Lemmas	Identical Expres.	Small Changes	Part of Idiom. Exp.	Free Translation	
Notices	48	37	3	0	8	16.7%
AboutThisBook	81	63	3	3	12	14.8%
Chapter 1	70	65	0	1	4	5.7%
Chapter 2	173	157	1	1	14	8.1%
Chapter 3	108	103	3	0	2	1.9%
Chapter 4	105	95	2	3	5	4.8%
Chapter 5	40	38	0	0	2	5.0%
Chapter 6	47	44	0	2	1	2.1%
Chapter 7	88	82	1	1	4	4.5%
Chapter 8	104	95	1	2	6	5.8%
Chapter 9	151	124	7	7	13	8.6%
Average						7.1%

	Verb					
	Number of Words					Ratio of Free Translation
	Repeated Lemmas	Identical Expres.	Small Changes	Part of Idiom. Exp.	Free Translation	
Notices	14	6	0	2	6	42.9%
AboutThisBook	21	7	2	2	10	47.6%
Chapter 1	26	13	2	2	9	34.6%
Chapter 2	50	20	5	5	20	40.0%
Chapter 3	30	21	3	1	5	16.7%
Chapter 4	39	21	6	3	9	23.1%
Chapter 5	9	4	1	0	4	44.4%
Chapter 6	17	13	2	0	2	11.8%
Chapter 7	22	9	0	1	12	54.5%
Chapter 8	28	13	6	1	8	28.6%
Chapter 9	54	22	3	4	25	46.3%
Average						35.5%

By sharing folders, you can... → Sharing folders allows you to...

Sometimes, predicates were expressed in both verb and noun forms. For example:

The attachment of A *to ...* → A *can also attach to ...*

> *The database supports deleting and reorganizing records.*
> *→ Database support also allows a record-level file definition.*

Such changes of expressions for predicates may improve the readability of the text; however, the necessity of implementing such a mechanism in an MT system depends strongly on the purpose of the system. Current MT systems have not yet reached a level at which free translation can be considered, and therefore we refrain from any further discussion of the matter here.

The results of this analysis also indicate some limitations of the approaches that uses bilingual corpora as knowledge resources, since the considerable amount of free translation in bilingual corpora may cause difficulties in matching corresponding expressions in two languages.

5 Conclusions

We have described the result of our analysis of computer manuals to show that there is actually a strong tendency for morphologically identical words to be used in the same way, and that a chapter consisting of around 2000 words is the most appropriate scope for the application of the discourse constraint in order to solve various problems in sentence analysis.

We have been developing a practical method of context processing by focusing on extracting inter-sentential information from the syntactic relations of morphologically identical words in different sentences in the same text. Our preliminary analysis shows that the frequency of morphologically identical words found in computer manuals is remarkably high (on average, in 12 computer manuals, 91.6% of all content words can be found in other sentences in the same manual, and 78.0% of all content words can be found four times or more in other sentences in the same manual), and that those words are usually repeated within a small area of a text (in more than 90% of all cases, morphologically identical nouns were found within 36 successive sentences, and in more than 72% of cases, morphologically identical nouns were found within 10 successive sentences). The results of the analysis in this paper, along with the results of our preliminary analysis, offer good prospects for the success of our approach to improving the accuracy of sentence analysis.

The results of another analysis that we made indicate that generating a good translation in a target language also involves the issue free translation. However, free translation may conflict with the preservation of consistency in translated expressions, which is widely considered to be a good feature of practical MT systems currently on the market. This issue may lead to a discussion on the direction of MT.

References

[1] J. Binot and K. Jensen: "A Semantic Expert Using an Online Standard Dictionary." In Proceedings of IJICAI-87 (1987)

[2] P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer: "Word Sense Disambiguation Using Statistical Methods,'" IBM Research Report (1991)

[3] W. Gale, K. Church, and D. Yarowsky: "Using Bilingual Materials to Develop Word Sense Disambiguation Methods," In Proceedings of TMI-92 (1992)

[4] J. Hobbs: "Resolving Pronoun References," Lingua 44 (1978)

[5] K. Jensen: "PEG: The PLNLP English Grammar," In Natural Language Processing: The PLNLP Approach, K. Jensen, G. Heidorn, and S. Richardson, eds., Boston, Mass.: Kluwer Academic Publishers (1992)

[6] V. Sadler: "The Textual Knowledge Bank: Design. Construction, Applications," In Proceedings of International Workshop on Fundamental Research for the Future Generation of NLP (1991)

[7] C. Sidner: "Focusing in the Comprehension of Definite Anaphora," In Computational Models of Discourse, M. Brady and R. Berwick, eds., Cambridge, Mass.: MIT Press (1983)

[8] N. Uramoto: "Lexical and Structural Disambiguation Using an Example-Base," In Proceedings of the 2nd Japan-Australia Joint Symposium on Natural Language Processing (1991)

[9] T. Utsuro, Y. Matsumoto, and M. Nagao: "Lexical Knowledge Acquisition from Bilingual Corpora," In Proceedings of COLING-92 (1992)

[10] "IBM SAA ImagePlus Object Distribution Manager MVS/ESA High-Speed Capture Subsystem Guide Version 2 Release 1.1," IBM Corp. (1991)

[11] "IBM Application System/400 New User's Guide Version 2," IBM Corp. (1992)

[12] "Longman Dictionary of Contemporary English," Harlow and London: Longman Group Limited (1978)

```
"Chapter 6 ."
"第6章"
((NP   (NOUN*  "chapter" ("chapter" SG))
       (LABEL  (NOUN*  "6" ("6" PL)))
       (PUNC   "."))  0)

"Storing Messages in Journal Files ."
"ジャーナル・ファイルへのメッセージの記録"
((PRPRTCL(VERB*"storing" ("store" PG))
         (NP      (NOUN*  "Messages" ("message" PL))
          ?       (PP     (PP     (PREP*  "in"))
                          (NP     (NOUN*  "Journal" ("journal" SG)))
                          (NOUN*  "Files" ("file" PL))))
       (PUNC   "."))  500.6458125)
```

Figure 1: Example of the Contents of a Text Database

```
number    number
          (Sentence 88 of CHAPTER4 )
          "nnnn is the task number for DB2 ."
          "nnnn  DB2のタスク番号。"
number    number
          (Sentence 143 of CHAPTER4 )
          "tally1 is the number of ImagePlus workstations successfully initialized ."
          "tally1   正しく初期設定されたImagePlusワークステーションの数"
number    number
          (Sentence 144 of CHAPTER4 )
          "tally2 is the number of ImagePlus workstations that were not initialized ."
          "tally2   初期設定されなかったImagePlusワークステーションの数"
number    number
          (Sentence 147 of CHAPTER4 )
          "The value that appears in place of tally2 indicates the number of ImagePlus
           workstations that were not initialized and cannot be involved in image
           operations at this time ."
          "tally2に示される値は、初期設定されていないため現在のイメージ操作で使用できない
           ImagePlusワークステーションの数を示します。"
number    number
          (Sentence 209 of CHAPTER4 )
          "You must convert this number to the negative decimal DB2 code ."
          "この数字は、負の10進DB2コードに変換しなければなりません。"
```

Figure 2: Data on *number* in Chapter 4