ORIGINAL RESEARCH

Analysis on descriptions of precautionary statements in package inserts of medicines

Keita Nabeta[1]
Masaomi Kimura[2]
Michiko Ohkura[2]
Fumito Tsuchiya[3]

[1]Graduate School of Engineering and Science, Shibaura Institute of Technology, Toyosu 3-7-5, Koto-ku, Tokyo, 135-8548 Japan; [2]Faculty of Engineering, Shibaura Institute of Technology, Toyosu 3-7-5, Koto-ku, Tokyo, 135-8548 Japan; [3]School of Pharmacy, International University of Health and Welfare, Minami-Aoyama 1-24-1, Minato-ku, Tokyo, 107-0062 Japan

Correspondence: Keita Nabeta
Toyosu 3-7-5, Koto-ku, Tokyo,
135-8548 Japan
Tel +81 3 5859 8507
Fax +81 3 5859 8507
Email m709102@shibaura-it.ac.jp

Background: To prevent medical accidents, users must be informed of the cautions written in medical package inserts. To realize countermeasures by utilizing information systems, we must also implement a drug information database. However, this is not easy to develop, since the descriptions in package inserts are too complex and their information poorly structured. It is necessary to analyze package insert information and propose a data structure.

Methods: We analyzed the descriptions of 'precautions for application' in package inserts via text mining methods. In order to summarize statements, we applied dependency analysis to statements and visualized their relations between predicate words and other words. Furthermore, we extracted words representing timing to execute the order.

Results: We found that there are four types of statements: direct orders such as "使用する" (use), causative orders such as "使用させる" (make someone use), direct interdictions such as "使用しない" (do not use), and causative interdictions such as "使用させない" (do not make user use). As for words representing timing, we extracted six groups: "at the time of delivery," "at the time of preparation," "in use," "after use," and "at the time of storage." From these results, we obtained points of consideration concerning the subjects of orders in the statements and timing of their execution.

Conclusion: From the obtained knowledge, we can define the information structure used to describe the precautionary statement. It should contain information such as the actions described in the statement, the flag to express an order or interdiction, the subject to be ordered, and the timing.

Keywords: medical safety, drug information, package insert, drug database, text mining

Introduction

To prevent medical accidents, users must be informed of the cautions written on the package inserts of medicines. These are exclusive legal documents that describe detailed information for each drug, such as the composition, efficacy, dosage, and cautions. However, since the descriptions in package inserts are excessive and there are over 20,000 medicines in Japan, it is not easy for healthcare workers to read all documents.

As a countermeasure, the utilization of information technology has been focused on in recent years. To develop an information system that alerts healthcare workers at the user interface of an ordering system and an electronic health record system, we need to implement a drug information database as a source of information alerts. In order to utilize the same in information systems, the Pharmaceutical and Medical Devices Agency (PMDA)[1] has provided Standard General Markup Language

19

(SGML) formatted data of package insert information since the year 1999.[2] Some medical institutions have developed drug databases based on the SGML data to retrieve drug information.[3] However, this is not easy since the descriptions in package inserts are too complex and their information poorly structured.[4–8] In this SGML data, most information is directly stored as statements and terms have not been standardized or encoded. Since computer systems cannot interpret drug information, existing systems cannot check prescriptions and administration based on a drug information database. In order to develop such systems, it is necessary to develop a drug information database which has a structure suitable for the purpose.

In this paper, we report on the results of analysis concerning the descriptions of 'precautions for application' in package inserts in terms of meaning via text mining methods. The reason for analyzing 'precautions for application' data is that it contains information concerning the usage situation that must be known at the time of prescription. We also propose their data structure to organize the information in the descriptions.

Target data

We obtained the SGML data from the PMDA Web site, utilizing the "YJ code" as a key, which is a drug identification code and included in the "standard drug master (9/30/2007)" provided by The Medical Information System Development Center (MEDIS-DC).[9] There were 11,685 SGML files.

Preprocessing

In our target SGML data, since precautionary statements are described in the "precautionsforapplication" field (as shown in Figure 1), we extracted text data from the "item" and "detail" fields in the same. After this extraction, we split each sentence and converted 1 byte characters to 2 bytes. In this study, we removed the sentences described in parentheses, since these can be regarded as additional information. As a result, we obtained 43,639 statements from 9,130 SGML data.
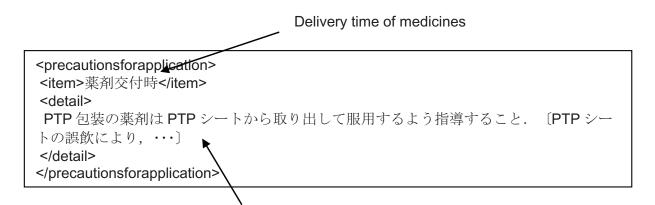
Additionally, we applied dependency analysis to the extracted statements via software named CaboCha (v0.53; Nara Institute of Science and Technology, Nara, Japan).[10] This technique extracts words, parts of speech, segments, and dependency relations between each segment by probabilistic calculation based on defined grammar and dictionary data. In this study, we investigated them by using information such as morphemes, segments, dependency relations, and parts of speech provided by the software.

Analysis
Extraction of predicate words

Since predicate words are often described at the end of the sentence in Japanese, we extracted only headwords in the final segment from the results of dependency analysis. Table 1 shows the top ten extracted words and their frequencies.

This result shows that "こと" (must) was the most commonly occurring word, with a frequency of 83%. It emerges with meanings of "order," if we place it after verbs, eg, "使用すること" (must use). We therefore focused on verbs that modify "こと" (must), although there are other order expressions that are weaker, eg, "望ましい" (desirable) and that of possibility "できる" (can).

Delivery time of medicines

```
<precautionsforapplication>
 <item>薬剤交付時</item>
 <detail>
  PTP 包装の薬剤は PTP シートから取り出して服用するよう指導すること．〔PTP シートの誤飲により，・・・〕
 </detail>
</precautionsforapplication>
```

Guide to pick up a medicine packed in a PTP sheet. [By accidental ingestion…]

Figure 1 Sample data of precautionary statement.

Table I Extracted predicate words (top 10)

Predicate words	Frequency	Ratio (%)
こと (must)	24,929	82.86
望ましい (desirable)	1578	5.24
ある (exist)	812	2.70
洗い流すこと (must wash)	250	0.83
注意 (caution)	194	0.64
場合 (case)	105	0.35
できる (can)	73	0.24
粘膜 (mucous membrane)	68	0.23
行う (do)	51	0.17
開瞼する (keep the eyelids apart)	47	0.16

Extraction of order statements

To extract information concerning the order description, we collected verbs modifying "こと" (must) in the previous subsection and obtained 415 kinds of headwords in Table 2. In this table, we can see that there are verbs expressing direct orders such as "使用する" (use), causative orders such as "使用させる" (make someone use), direct interdictions such as "使用しない" (do not use), and causative interdictions such as "使用させない" (do not make user use).

Since the descriptions of order and interdiction coexist in precautionary statements, medical staff can wrongly receive such information. Therefore, it is necessary to make order/interdiction flags in the database and to provide medical staff with the flags separately by information systems. Furthermore, since direct and causative statements coexist, targets to be ordered are not clear. Although humans can read them and understand targets, it is difficult to make computers interpret targets. In order to alert to appropriate targets, it is necessary to identify the appropriate targets to be alerted.

Focusing on these points, we categorized the statements into four categories: direct orders, causative orders, direct interdictions and causative interdictions. In order to summarize information for each category, we tried to

Table 2 Extracted order words (top 10)

Order words	Frequency	Ratio (%)
指導する (guide)	4771	18.27
注意する (be careful)	3696	14.15
使用する (use)	3417	13.08
使用するない (do not use)	2440	9.34
注射する (inject)	986	3.78
避ける (avoid)	968	3.71
投与する (administer)	918	3.52
行う (perform)	760	2.91
行うない (do not perform)	436	1.67
する (do)	420	1.61

employ the Word-Linking Method.[11] However, we cannot obtain relations between verbs and subjects/objects using this technique because the method extracts only nouns, verbs, and adjectives.

To extract such relations, we focused on the case markers (particles) appearing at the end of segments that modify the verbs. These are words that indicate the dependency relationship of a certain word to a predicate verb. Although we obtained three types of case markers using CaboCha: "case markers – general," "case markers – collocation," and "case markers – citation," we used only case markers – general, since it accounted for 90.65% and there were fewer case markers in that type than others. In this study, we focus on the case markers, "を" (wo), "に" (ni), "が" (ga), "から" (kara), "で" (de), "と" (to), "へ" (e), "の" (no), "より" (yori), and "にて" (nite).

We aggregated ternaries, a verb, a case marker, and the verb modifier, which also appears in the segment containing the case marker. After aggregation, we visualized the ternaries as nodes of a network to find major information in the order statement concerning the four cases, which are direct order, interdiction, causative order, and causative interdiction. In the network, the nodes are collected verbs, case markers, and keywords, and the edges express the connections among them, while their frequencies are expressed as their width. The threshold of appearance frequencies in the data limits the nodes and edges in Figures 2–5.

Direct order statements

Figure 2 shows that major direct order verbs are "使用する" (use), "注意する" (be careful), and "中止する" (stop); we can also see statements such as "点眼用に使用する" (use for eye drops), "静脈内に注射する" (inject into vein), and "洗い落としてから使用する" (use after washing).

It should be noted that the statement "神経走行部位を避けること" (avoid part of nerves) appeared in this network. Though it is an order to avoid, it can be interpreted as an interdiction, since the avoidance is implied in the concept of interdiction.

Direct interdiction statements

Figure 3 shows that the major expressions concern the interdiction of usage. In particular, we can see an interdiction to apply medicines to the affected areas, such as "炎症部位に使用しない" (do not apply to inflammatory part). Moreover, we obtained statements that indicate an interdiction of mixing drugs, eg, "製剤と混合しない" (do not mix with drugs).
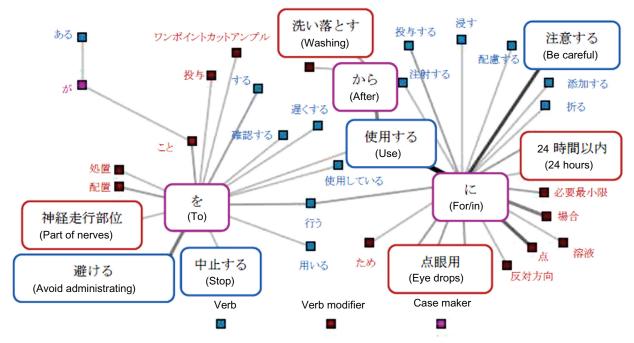
Figure 2 Network of words related to direct orders (threshold: 90).

Causative order statements

In Figure 4, we see two types of causative verbs. The first is for medical experts and patients, and another type is for medicines.

The former gives examples such as "噛まずに服用させる" (make patients take the medicine without chewing), "食後に服用させる" (make patients take the drug after a meal), and "水で服用させる" (make patients take the medicine with water). These are statements that indicate orders to patients via pharmacists and nurses.

The latter gives examples such as "中和させる" (neutralize) and "乾燥させる" (dry). They express methods for the preparation and storage of medicines.

Causative interdiction statements

As causative order statements already mentioned, the subjects of the former include both humans (pharmacists, nurses and patients) and medicines. In Figure 5, statements such as "水なしで服用させない" (do not make patients avoid
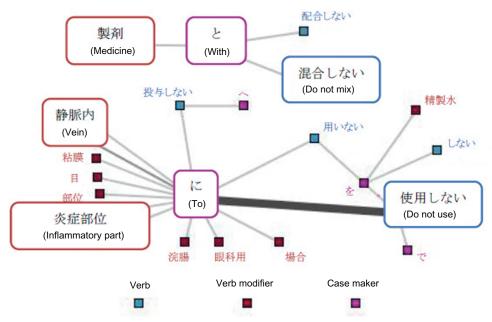

Figure 3 Network of words related to direct interdiction (threshold: 50).
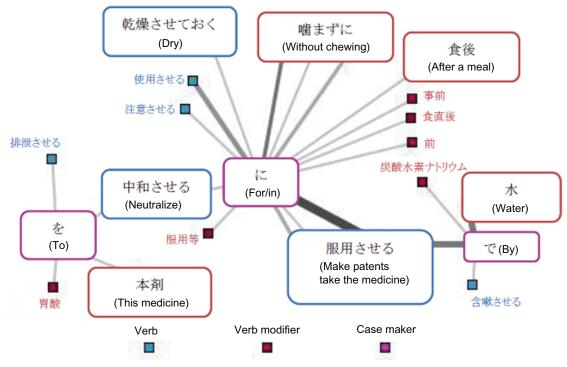
Figure 4 Network of words related to indirect orders (threshold: 5).

water) and "注射筒から逆流させない" (do not regurgitate from syringe) can be found.

Based on these results, we found two points concerning precautionary statements that should be considered. The first is that both order statements and interdiction statements chaotically coexist. Since this may lead to the misreading

of these statements, they should be described separately. The second point is that we should distinguish two kinds of subjects of statements, namely patients and medical experts. This is important for system checking because the system must know the subjects to whom warnings should be given.
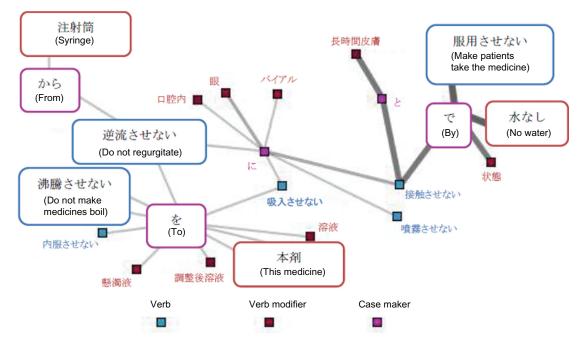

Figure 5 Network of words related to indirect interdiction (threshold: 0).

Table 3 Extracted adverbial nouns (top 10)

Adverbial nouns	Frequency	Ratio (%)
場合 (case)	2336	25.13
刺したとき (time of stinging)	599	6.44
ため (because)	536	5.77
後 (after)	357	3.84
開封後 (after opening)	346	3.72
とき (time)	303	3.26
長期間 (long-term)	235	2.53
長時間 (extended period)	218	2.34
アンプルカット時 (time of ampoule cutting)	160	1.72
溶解後 (after dissolving)	153	1.65

Extraction of words representing timing

In the previous section, we found words that express the time to perform an action eg, "投与前" (before administration) and "使用時" (in use). Orders should be provided to medical staff and patients at the appropriate time. In order to clarify correspondences between timing and orders, we extracted words representing timing from the target data and classified them based on procedures to use drugs.

Initially, we compiled keywords where the part of speech is an adverbial noun (Table 3). We defined "場合" (case), "前" (before), "後" (after), "時"/"とき" (time), and "際" (when) as feature words that express the concept of timing, since most segments expressing timings contain these words.

Next, we extracted expressions of timing via the characters defined above. The precautionary statements contain mainly two types of descriptions, one of which is that the order is described after the timing expressions in statements eg, "使用時に注意すること" (be careful at the time of use), and another of which where the timing expression appears singly eg, "薬剤交付時" (delivery time of medicines) in the "<item>" field of SGML data. In the former case, we extracted keywords that modify the verb expressing the order

and whose segments contain the feature words defined above. In the latter case, we extracted keywords that are at the tail of sentences and have feature words. We regarded the keywords obtained as timing to execute the order.

The resultant keywords contained those with similar meanings. To reduce the variety of expressions, we identified such synonymous expressions by hands. We investigated the top 50 extracted words and found six groups: "at the time of delivery," "at the time of preparation," "in use," "after use," and "at the time of storage." We show the results of this classification in Table 4.

We found that the caution related to "in use" was the largest group. Utilizing this classification, we obtained characteristic words that express an order and appear only at specific instances, eg, "廃棄する" (dispose of) after use and "遮光保存する" (store under light interception) at times of storage. The obtained timing for alerts are vital to prevent accidents in order for the system to specify the time at which to send an alert.

Conclusion

In this study, we analyzed precautionary statements in the package inserts of medicines using a text-mining method to develop a drug information database. Consequently, we found the major statements used to express orders and interdictions in use. Moreover, we obtained points of consideration concerning the targets of orders in the statements and timing of their execution. In this study, although we applied several methods to text data written in Japanese, our methods can be employed for any language if a dependency analysis is applied to them. We can also use prepositions in English or other languages instead of particles in Japanese language. This suggests that our method is applicable to languages other than Japanese.

Utilizing the knowledge obtained, we can define the information structure used to describe the precautionary statement. It should contain information such as the actions

Table 4 Classification of words meaning time

Class	Example of extracted words	Frequency	Ratio (%)
Time of delivery	薬剤交付時 (delivery time of medicines)	3791	13.77
Time of preparation	調製時 (time of preparation), アンプルカット時 (time of cutting ampoule), etc	6067	22.04
In use	投与時 (in administration), 注射時 (time of injection), etc	12,599	45.77
After use	投与後 (after administration), 使用後 (after use)	279	1.01
Time of storage	保存時 (time of storage)	406	1.48
Others	やむを得ない場合 (in the inevitable case), 反復使用する場合 (for the repetitive use), etc	1438	5.23

24

Drug, Healthcare and Patient Safety 2012:4

described in the statement, the flag to express an order or interdiction, the subject to be ordered, and the timing. In the future, this database will be accessed from various systems such as ordering systems and portable information terminals for nurses and patients, and these systems will provide orders and cautions to the appropriate targets at the appropriate time. Namely, the systems will provide orders of administration methods to patients and/or nurses just before administration.

In future work, we will analyze other information in the package inserts and develop a drug information database based on the results.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Menu of package insert information: Pharmaceuticals and medical devices agency; [cited October 15, 2007]. Available from: http://www.pmda.go.jp/. Accessed September 6, 2011.
2. Orii T, Yamazaki M. Utilization of drug information through the internet. *Japan Journal of Medical Informatics*. 2001;20(6):469–486.
3. Fujimoto K, Kuwata S, Matsumura Y, et al. Constructing of drug information system on web with utilizing pharmaceuticals information providing system. *Japan Journal of Medical Informatics*. 2002;22(4): 333–340.
4. Otsuki C, Ota T, Sunagane N, et al. Study on inconsistency in epileptic seizure terms used in package inserts. *Japanese Journal of Pharmaceutical Health Care and Sciences*. 2005;31(1):65–71.
5. Togashi H, Kuribara M, Orii T. Analysis for contents data of package leaflet in medicinal supplies. *Japan Journal of Medical Informatics*. 2006;26(2):129–134.
6. Hamada M, Hirota M, Kurata K, Dobashi A. Creation of drug information database based on pharmaceutical markup language (PML). *Japanese Journal of Pharmaceutical Health Care and Sciences*. 2007;33(6):502–509.
7. Nabeta K, Kimura M, Ohkura M, Tsuchiya F. Investigation into problems of utilization of drug information in package inserts to ensure the safety of drug usage. Proceedings of Applied Human Factor and Ergonomics International 2008. July 14–17, 2008; Las Vegas, United States.
8. Kimura M, Okada K, Nabeta K, Ohkura M, Tsuchiya F. Analysis on descriptions of dosage regimens in package inserts of medicines. *Lect Notes Comput Sci*. 2009;5618:539–548.
9. HOT code master: The Medical Information System Development Center; [cited October 30, 2007]. Available from: http://www.medis.or.jp/. Accessed September 6, 2011.
10. Yet Another Japanese Dependency Structure Analyzer: Kudo T; [cited October 30, 2007]. Available from: http://code.google.com/p/cabocha/. Accessed September 30, 2011.
11. Kimura M, Furukawa H, Tsukamoto H, et al. The analysis of questionnaires about safety of drug use, the application of text mining to free description questionnaires. *Japanese Journal of Ergonomics*. 2005;41(5):297–305.