

Business Collaboration Support for Offshore Software Development

Takashi Moriyasu[†]

Advanced IT Laboratory

Toshiba Solutions Corporation Fuchu, Tokyo 183-8512, JAPAN

Tel: +81-42-340-6400, E-mail: Moriyasu.Takashi@toshiba-sol.co.jp

Guowei Zu

Advanced IT Laboratory

Toshiba Solutions Corporation Fuchu, Tokyo 183-8512, JAPAN

Tel: +81-42-340-6400, E-mail: So.Kokui@toshiba-sol.co.jp

Hiroshi Tsuji

Department of Computer Science and Intelligent Systems, Graduate School of Engineering,

Osaka Prefecture University Sakai, Osaka 599-0581 JAPAN

Tel: +81-72-254-9353, E-mail: tsuji@cs.osakafu-u.ac.jp

Received, March 30, 2010; Revised, July 16, 2010; Accepted, July 16, 2010

Abstract. Offshore software development (OSD) is international business collaboration. OSD projects often encounter intercultural and inter-linguistic problems disturbing the projects. Business documents are formal media of information and knowledge for OSD. While OSD documents should convey common understanding of the OSD products, the documents may contain unsuitable expressions which draw misunderstanding of the required products and offensive issues for the collaboration. Intercultural and inter-linguistic differences cause mistakes and inappropriate expressions. OSD from Japan to China is the largest in Asia, and Japanese language is often used in OSD documents. Large similarity is found between Japanese and Chinese in their languages, while many differences exist even for the same word. The similarity induces to write unsuitable expressions for both sides of OSD. To introduce risks for OSD projects caused by unsuitable or inappropriate expressions in OSD documents, we propose to apply a proofreading system of Japanese documents for OSD. Japanese consignor uses the system to refine OSD documents written by Japanese engineers for Chinese readers, and Chinese consignee uses it to refine Japanese documents written by Chinese Engineers as derivatives of OSD projects. Effectiveness of applying the proofreading system is discussed for actual projects.

Keywords: Offshore Software Development, Document Proofreading, Cultural and Linguistic Differences

1. INTRODUCTION

Globalization of business brings risks and problems for business collaboration. Collaboration with partners in foreign countries implies risks and problems caused from business interchange over different cultures, customs, and often languages. Offshore software development (OSD) is typical international business collaboration, and meets with those risks and problems.

Offshoring is outsourcing of a business process by a company in a country to other companies usually located in other countries. In computer software engineer-

ing industry, OSD rapidly grows its amount along with the globalization of business and development of information and communication technologies (Friedman 2006). OSD is intelligent and highly skillful collaborative work.

English is the common language for business collaboration in the flat world. India is the largest recipient country of OSD outsourcing in the world. Capability of English language of Indian software engineering companies and their engineers plays a large part in achieving the top market share in English speaking countries including the U.S., the largest market of OSD.

In Asian region, OSD from Japan to China is the

[†] : Corresponding Author

largest amount. In these decades, Japanese IT companies have been significantly expanding their OSD, especially to China. MIC (2007) reported that the main offshoring destination country of Japan is China (79.2%), and that of the US is India (94.3%). China has made a great advancement in building infrastructure for software engineering industry, such as construction of many software parks. Chinese government has also expressed the clear posture to support business deals with Japan.

Japanese language capability of OSD recipient companies in China is also a major factor of OSD growth from Japan. MIC (2007) reported that the largest factor for Japanese companies to choose country of OSD recipient is low offshore price (85.4%) and the second factor is availability of workers with Japanese ability (68.8%), and that the largest factor for Japanese companies to choose OSD recipient companies is availability of workers with Japanese ability (70.8%), the second is low offshore price (59.4%), and the third is availability of highly skilled workers (54.2%). Japanese language capability of Chinese software engineering companies and of their engineers promotes much OSD from Japan to China.

Long history of cultural exchange between Japan and China generated a lot of similarity in their cultures and languages. However more differences exist of course. Differences hidden behind the similarity are sometimes passed by Japanese people and also by Chinese. It may cause critical risks to cause problems for business collaboration and for OSD.

In this paper, we discuss on risks and problems of OSD as business collaboration, and especially of communication for business collaboration. We propose support for OSD and international business collaboration by improving document quality with using proofreading technology in order to reduce such risks and problems on OSD documents.

In the following section we mention risks for OSD projects. Focusing on communication for international business collaboration, the Section 3 discusses risks and problems on business documents for OSD. In Section 4, we propose document processing support for business collaboration in OSD to diminish the risks, and we discuss on result of our approach in Section 5.

2. RISKS FOR OFFSHORE SOFTWARE DEVELOPMENT PROJECTS

2.1 Development of Japanese OSD and its Risks

Tsuji *et al.* (2008a) observed OSD evolution stages and defined the stages concerning with CMMI levels. They described as follows.

In the first stage, OSD projects are sorts of experimental dispatch of software development to countries with low personnel cost. Many OSD projects then suf-

fered unsuccessful results.

In the second stage, learning from failures in the first stage, OSD consignors make more of high skill human resource availability, and process management and quality management are introduced in recipient companies.

Successful experiences evolve OSD into the third stage where consignor and consignee regard each other as partners and recognize OSD as business collaboration.

Many Japanese companies are now conceived as in the third stage through OSD experiences in more than several year practices. They still suffer from risks and problems of OSD even in the third stage.

OSD contains complex of risks caused from both aspects of software engineering and international collaborations. International business interchange implies critical risks over different cultures and customs often with language barrier, and they complicate fundamental risks of software engineering.

While numbers of managers and engineers have experienced the success and failures of their OSD projects, their knowledge and know-how skills remain as tacit knowledge. Tsuji *et al.* (2007b) proposed questionnaire-based risk assessment scheme for Japanese OSD projects. They developed a method to quantify the risk of OSD projects by designing questionnaires about OSD project evaluation with fourteen attributes related OSD. In addition, Tsuji *et al.* (2008b) defined a systems engineering approach to extract explicit knowledge from tacit one for OSD. Wada *et al.* (2008) proposed a method for critical factor analysis of OSD success by structural equation modeling. Sheng *et al.* (2008a) proposed an experimental risk estimation method based on questionnaire for externalizing the know-how of experienced project managers. Tsuji *et al.* (2007a) proposed conjoint analysis to detect risk factors on OSD, and revealed the following inclination;

- Changes of requirement specification are the largest factor to OSD project success/failure in the aspect of product software, and communication capability is the largest factor in the aspect of selection of recipient vendors. Cost reduction is still the largest motivation for OSD. They also distilled difference of critical factors for vendors between India and China; Project management capability is the most important for Indian companies, and communication capability is the most for Chinese companies. Saito (2007) discussed importance of communication management for OSD from Japan to China.

Ye *et al.* (2007) discussed importance of communication and coordination in OSD, and proposed the Dynamic Community framework to reduce the overall cost of communication and coordination in OSD. Christiansen (2007) focused on communication in OSD projects, and described that communication is important to achieve all the advantages of OSD, such as access to talent, greater flexibility in access to resources and cheaper resources, and that face to face communication may be

impossible because the team members of the OSD project are located in different parts of the world, making different time zones.

2.2 Communication media for OSD Projects

Synchronous and asynchronous communication channels are available for OSD project members even in distributed in different countries by using satellite communication lines and through internet. Real meetings, telephone calls, video conferences, desktop conference systems over the internet, and chat systems are synchronous communication. E-mail is a typical asynchronous communication. Document circulation mostly through internet is also an asynchronous and classic one for business collaboration.

Business documents are the formal media of communication for business collaboration. Business documents are fundamentals for information and knowledge sharing in business collaboration. Software development is a very complicated discipline, and requires ample and conscientious communication among members of collaboration team, even located in different countries sometimes in different time zones, such as for an OSD project.

Various sorts of documents are exchanged between both shores for an OSD project. In the starting phase of the OSD project, documents are sent from the consignor to the consignee, such as specifications of the software products, requirements for the products and developing process, procedures about project management, and so on. The consignee sends to the consignor also volumes of documents such as project progress reports, detail design specifications, programming specifications, test

reports, manuals on installation, integration, operation, and maintenance, and so forth.

The following section, we discuss risks and problems which may be carried in business documents for OSD.

3. PROBLEMS OF DOCUMENTS FOR OSD

While documents for OSD should convey common understanding of the OSD project and its products, they may include risks drawing misunderstanding of the required products or offensive issues on the project. Documents sent from the consignee also often contain unsuitable expressions on the project and the product.

OSD from Japan to China is the largest amount in Asian region and Japanese is a formal language for OSD documents. Many Chinese OSD companies accept Japanese documents for OSD, and many of them can deliver Japanese documents to Japanese customers. Chinese software engineers then need read documents written in Japanese, such as the requirements and specifications for the target software, by themselves or sometime with the help of translators. They are also required to write documents in Japanese by themselves or with the help of translators. The documents are to deliver to their Japanese customers.

A number of Chinese people including software engineers study Japanese as a foreign language and can read and write Japanese. Japanese language capability certification is sometimes required to work for those OSD companies. The following two issues are on OSD document processing (Figure 1):

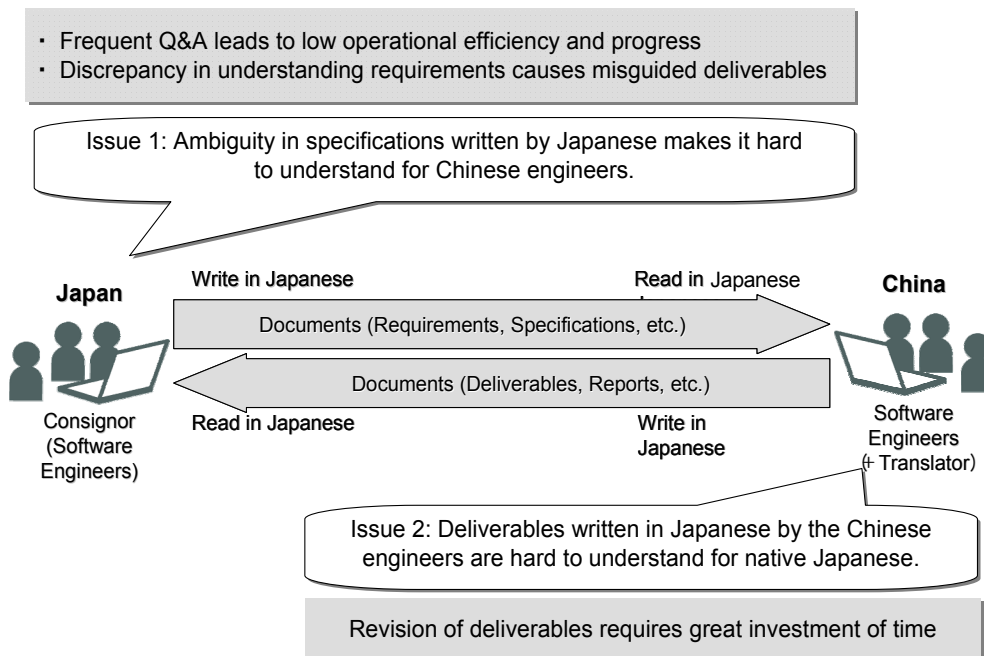


Figure 1. Typical problems of business documents in the OSD originated from Japan to China.

- (1) Chinese software engineers are confronted with likely a troublesome task to read and understand documents written in Japanese, such as requirements, and design specifications of target software.
- (2) Japanese consignors likely need delicate review of documents even in Japanese written by the Chinese engineers.

It is challenge for the Chinese engineers, even qualified of Japanese language capability, to completely realize what the specification means. It is also harder challenge to write adequate Japanese documents for Japanese customers.

We will investigate on what kinds of risks and problems emerge in the OSD originated from Japan to China in those situations.

3.1 Problems of Documents in Japanese Written by Japanese for Foreign Engineers

The following problems are found in OSD documents written in Japanese language by Japanese engineers of consignor companies.

3.1.1 Problems Caused from Linguistic Feature

Japanese expressions are hard to understand by linguistic features (Zu *et al.*, 2007b). Chinese characters are used abundantly in some countries other than China. Japanese language applies 1945 Chinese characters for usually use.

- (1) The meaning of the same expression between Japanese and Chinese however could be different meaning between both. For example, “8以下(*i-ka*)” in Japanese the expression means “8 or less”, but it means “less than 8” in Chinese. Also, the word “空色(*sora-iro*)” means “sky-color”; sky-blue in Japanese, but it means “empty-color”; colorless or white in Chinese. The difference on the same expression may cause critical misunderstanding which draws false development of the product software.
- (2) Katakana of Japanese is syllabary and is most often used for transcription of words from foreign languages. Technical documents abundantly contain variety of foreign words mostly from English. Even though those katakana words spelled like the original foreign words, the pronunciations are different from the original. They are difficult to be recognized their original words by non-native people, even who are native speaker of the origin words. Moreover, multiple katakana words are sc-ripped continuously, without any space or separator in between. “ファンクションボタンエリア (*fan* (3 letters)-*ku-sho*(2 letters)-*n-bo-ta-n-e-ri-a*)” for example means “function button area.”
- (3) Abbreviated katakana terms are another difficulty for foreigners. “アプリ(*a-pu-ri*)” is the abbreviation of “アプリケーション(*a-pu-ri-ke-e-sho*(2 letters)-*n*)” meaning “application.” These expressions cannot be

found in an ordinary Japanese dictionary, and are much harder to understand for non natives.

3.1.2 Problems Caused from Ambiguous Expressions

Ambiguous expressions of Japanese are critical barriers for foreign readers, as Kojima *et al.* (2007) described. They pointed out the ambiguity of Japanese expressions are often based on tacit understanding common for most Japanese. Sheng *et al.* (2008b) surveyed to find out the main risk factors from the vendor’s viewpoint, and found that the problems occur because the statement of the consignor is ambiguous.

Japanese is a language which considers the listener’s feelings, and favors gentle, indirect and euphemistic expressions. On the other hand, since it leaves the understanding of so many ambiguous expressions to the listener, it is very difficult for foreigners to understand the expressions who don’t know the Japanese customs very well.

The expression “必要に応じて (*hitsuyou-ni ohjite*)” for example means “according to necessity”, although it is so ambiguous without any definition or information on “necessity.” The reader have to judge the necessity is “what”, “for whom”, or so.

Unfixed quantity expressions are found in Japanese documents. For examples, “すばらしい (*su-ba-ra-shi-i*)” meaning “wonderful”, “少し (*suko-shi*)” meaning “a little/a few”, and “ちょっと (*cho-t-to*)” meaning “just for a moment” are considered ambiguous and unsuitable for a documents such as software specifications.

A lot of ambiguous expressions in specification documents in Japanese give rise of frequent questions from consignee engineers. Answering those questions disturbs engineers of the consignor, and is time consuming for both sides of the OSD. Frequent Q&A decreases business efficiency and grows the project-progress worse. To avoid these matters, it is necessary to use clear expressions and avoid unsuitable, vague expressions.

Other than above unfavorable expressions, there are barriers for foreigner to understand documents in Japanese, such as a lot of abbreviations, spoken languages, dialects, polysemous words (with multiple meanings relating each other), proverbs, and the very complex honorific expression systems. Restrictions on the usage of these words and phrases are required for OSD documents.

3.1.3 Problems Caused from Cultural Difference

Misunderstandings of foreign documents are also caused from expressions with difference in cultures on both shores. The countries involved in OSD have different social systems, law organizations and cultural backgrounds. There are also many differences regarding history, politics, religions and sentimental issues among these countries. Much attention on those issues, especially on compliance, should be paid in the communication and in documents. Those issues bring not only unsuccessful result of the OSD project, but also cause larger problems or damages for the partnership, moreover

arise issues of international relationship between the two countries on the both shore.

3.2 Problems of Japanese Documents Written by Foreign Engineers

The mistakes in Japanese documents written by the foreigners vary from the mistakes Japanese make (Zu *et al.*, 2007b). Those mistakes are in wide variety that it is hard to understand the context. Figure 2 shows the main problems in the delivery statements written by foreign engineers. We discuss distinguish mistakes for foreigners to write Japanese.

3.2.1 Particles of Japanese

Japanese is one of agglutinative languages including Turkish, Korean, Malay, Indonesian, Hungarian, and others. Grammatical relations of agglutinative languages are shown by the particles. Chinese is an isolated language, in which such like Vietnamese, Thai, and Tibetan are classified. Then Chinese characters show the meaning of words and the word order shows the grammar. These grammatical differences between both languages frequently cause particle mistakes in the Japanese documents written by foreigners. For example, particles “が (*ga*)” and “を (*wo*)”, or “を (*wo*)” and “に (*ni*)” are misused frequently and the lack of particle “の (*no*)” are often seen. Such mistakes are hardly seen in documents written by Japanese.

3.2.2 Japanese Honorifics Expressions

Complex system of honorifics reflecting the hierarchical nature of Japanese society in history is very hard for foreigners to grasp and to use properly. Its misuses often found in documents even written by native Japanese.

3.2.3 Foreign Language Style Left in Japanese

When a foreigner writes a document in Japanese, it is very common, not to write a document directly in

Japanese, but first to write it in his/her mother language and then translate it into Japanese. The grammar, expressions and words of the mother language may often be left in the target documents even in Japanese.

Many Chinese characters are used in Japanese, and many of them are the same in Japanese, but some are different, such as simplified ones and created in Japan. It is not easy for Chinese and also for Japanese to judge a character in thousands ones is the same or not in both languages. Then Chinese characters used only in China are often left in Japanese documents. Chinese style words are also left.

3.2.4 Katakana Notation

Japanese katakana is hard to read as described in Section 3, and it is also hard to spell appropriately. Mistakes of katakana notations are often found documents written by foreigners.

Other than these, mistakes such as unsuitable tense, unsuitable omissions or punctuation marks also occur frequently. In order to reduce these mistakes, a lot of time and labor are needed for reviews by consignees and consignors.

4. DOCUMENT PROCESSING SUPPORT FOR BUSINESS COLLABORATION IN OSD

Machine translation is a supportive tool to overcome linguistic barrier of international communication. Chinese-to-Japanese and Japanese-to-Chinese machine translation have distinguishably developed to support reading and writing foreign language (Izuha *et al.*, 2007), however present quality of output sentences is not enough to directly adopt in formal documents. Some kinds of unnatural expressions can be seen in the machine translated documents. Proofreading support is necessary even if machine translation is applied to produce documents in foreign languages.

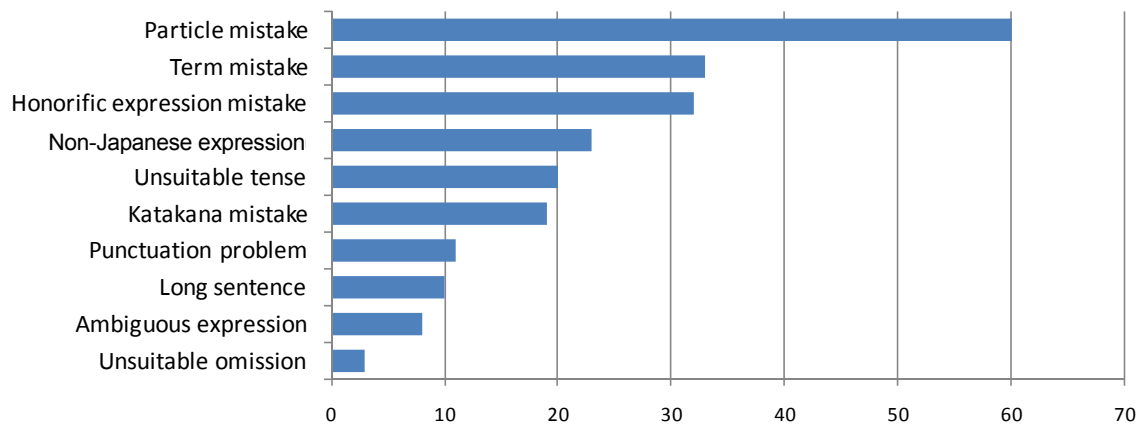


Figure 2. The main problems in the delivery statements written by foreign engineers.

4.1 Proofreading Support System for OSD Documents

Iwata (2005) proposed business document proofreading support in order to reduce compliance risks implied in documents. It detects mistakes and risks from the business documents based on related regulations, in-house rules, business knowledge, know-how, etc.

Using the supporting technology for business document proofreading, business documents can be checked on compliance, and also the document quality can be increased based on the business knowledge and know-how. This technology is applicable in various business fields, such as business document reviews, report check in medical field (Makino, 2005), financial and insurance document check, etc. Taniguchi *et al.* (2008) proposed a contextual proofreading system for numerical inconsistencies in business documents.

We propose to apply proofreading support systems for OSD document processing in two ways;

- (1) To support Japanese outsourcers to refine OSD documents in Japanese written by Japanese with quality and less risk of expressions. The output documents are also expected easy to read and to understand for Chinese consignee engineers.
- (2) To support Chinese and Japanese to refine OSD documents in Japanese written by Chinese.

The supporting technology for business document proofreading consists of natural language processing, automatic error detection, automatic document processing, and dictionary construction technologies. Proofreading engine consists of the following two subsystems;

- (1) Rule-based proofreading engine based on business knowledge and know-how.
- (2) Mistake detection using natural language processing.

Proofreading is closely connected to business knowledge and know-how. Characteristic expressions are extracted from false examples, related regulations, rules, customs, etc. Proofreading rules and dictionaries are built from those expressions. The rules and dictionaries are prepared for specific usage and/or organization, check documents according to specific tasks or business contents (Table 1).

Mistake detection is done by using natural language processing such as morphological and parsing analysis. Applying advanced check is planned by using coincidence, dependency, and case or semantic attribute information.

Business document proofreading needs consider foreign cultures. We are building check rules and dictionaries considering foreign cultures and customs based on sample data and know-how provided by experts working in proofreading fields, OSD ordering and management division, and OSD recipient companies.

The objectives of the proofreading of OSD documents are to detect following problems and unsuitable expressions automatically by using the natural language processing.

- (1) To detect unsuitable expressions.
- (2) To judge readability of a sentence.

Unsuitable expressions such as, ambiguous expressions in Japanese, expressions inviting risks, expressions easily misunderstood by foreigners, and so forth are collected and their patterns are registered into a dictionary. Detection of unsuitable expressions is executed by collation of patterns in the dictionary.

Readability is another criterion. For example, the longer the sentence becomes, the more difficult to understand it, especially for foreigners. Sentences with wrong punctuation marks are difficult to read. Readability metrics are introduced to judge readability of a sentence.

Table 1. Rules for the proofreading support system of Japanese documents written by Japanese for Chinese.

Issue	Check Item	Examples
Misleading Japanese expressions due to linguistic feature	Same word in Chinese with different meaning (homographs)	空色, 空白, 以下, 以上
	Long sentence	文字数, 単語数, 文節数が閾値を超える
	Double (triple or more) denial	～ないわけではない
	Unsuitable katakana word(s)	ファンクションボタンエリア, アプリ
Ambiguous expression	Demonstrative reference	これ, それ, あれ, 前述の, 後述の, こんな, このような, ある (日, 人, もの), その場合
	Simile	～ようである, ～らしい, みたいな, ～感じの
	Volume, grade, range, or intensity	多分, 大たい, 大抵, おおよそ, ほとんど, あまり, 非常に大きい, あたり, そのくらい
	Reference standard	最新に, 必要な場合, 必要に応じて, 必要以上の, 本来の
	Inarticulate phrase	いいです, 結構です
Expressions with cultural differences	Obscure phrase	ほうがいい, しなくてもよい, かもしれない
	Discriminatory phrase	(Omitted)
	Phrases on politics, religion, or faith	(Omitted)
	Term popular only in Japan	Japanese dating system, trade names, and proprietary terms

Instead of number of characters in a sentence, those features are applied such as number of words in a sentence, number of clauses from syntax analysis, number of characters between punctuation marks, number of Chinese characters, number of continuous katakana characters, etc.

We are focusing on documents processed in OSD from Japan to China in the following.

4.2 Support of Processing Japanese Documents for OSD

We propose a specification check system automatically checking expressions that are hard to understand

for Chinese engineers, or unsuitable from technical documents such like specifications, design documents and so on for OSD to China. It is based on the supporting technology for business document proofreading (Zu, 2007a). The check results are returned to the users in the form of alarms or comments attached in the original document.

The check of Japanese-written specifications is implemented, to make them intelligible for the Chinese engineers. The system is also applied to proofreading of delivery statements written by foreigners. Table 2 shows the rules for the proofreading support system of Japanese documents written by Japanese for Chinese.

Figure 3 shows a part of output of the proofreading

Table 2. Rules for the proofreading support system of Japanese documents written by Chinese.

Issue	Check Item	Examples
Complicating expressions	Same word in Chinese with different meaning (homographs)	空色, 空白, 以下, 以上
	Long sentence	文字数, 単語数, 文節数が閾値を超える
	Ambiguous expression	わりに, そちら, 前に, 比較的
Unsuitable expressions	Katakana error	コンディング(コーディング), アドレス(アドレス)
	Chinese-specific expression	等于(等しい), 宣講(講演), 中外(内外)
	Informal expression	こんな, それで, とか
Incoherent notations	Writing style	ですます体(調)と, である体の混用
	Mix of single/double byte characters	図2-2, API, API
	Format	Paragraph style, headings, citation, ...
Grammatical errors	Particle mistake	システムのコストを増える
	Lack of particle	改行追加する

4. 「メニュー」シートを開く。

5. 「tomcat 起動」 ボタンを押して、tomcat を起動する。

6. 「辞書リスト」更新ボタンを押して、プルダウンリストに最新の辞書リストが表示されるようにする。

7. 「データファイルリスト更新」ボタンを押して、プルダウンリストに最新のデータファイルリストが表示されるようにする。

8. 辞書編集が必要な場合は、「辞書名」プルダウンリストで該当辞書を選び、「辞書を編集」ボタンを押して辞書を開く。辞書編集後、Excel ファイルとして辞書を保存し、ファイルを閉じる。

9. 分析に使う辞書を「辞書名」プルダウンリストから、分析を行うデータを「データファイル名」プルダウンリストから選択し、「マイニング実行」ボタンを押して、テキストマイニングを開始する。コマンドプロンプトが停止したら、分析は終了である。

10. 「結果表示」ボタンを押し、分析結果をブラウザで確認する。

11. 必要に応じて、8~10 の操作を繰り返す。

12. テキストマイニング終了時は、結果表示ブラウザを閉じ、「tomcat 終了」ボタンを押して tomcat をシャットダウンしてから、テキストマイニング LiteV3.xls を閉じる。

2.3. テキストマイニング辞書作成
テキストマイニングの辞書は、文書中に現れる該当表現を「項目」に集約するための設定

コメント [s17]: 英語原語表現を併記してください。

コメント [s18]: 英語原語表現を併記してください。

コメント [s19]: 参照基準が曖昧です。具体的にどこを参照するか明記して下さい。

コメント [s20]: 英語原語表現を併記してください。

コメント [s21]: 参照基準が曖昧です。具体的にどこを参照するか明記して下さい。

コメント [s22]: 参照基準が曖昧です。具体的にどこを参照するか明記して下さい。

コメント [s23]: 英語原語表現を併記してください。

コメント [s24]: 長文です。各文は 80 文字以内にして下さい。

Figure 3. Output of the proofreading system for a sample of a specification for OSD written by Japanese.

system for a sample specification document for OSD written by Japanese. The proofreading system attaches comments over the original document. Each comment indicates a word or a sentence segment possibly unsuitable for a specification document of OSD. The comment reads recommended instruction to revise the expression.

The authors revise the document with referring to the comments. The revised document is reviewed by the manager or the project leader, who approves it to send to the consignee if it is appropriate.

5. DISCUSSION ON APPLYING SUPPORT SYSTEM FOR OSD DOCUMENT PROCESSING

The proposed support system is applied to real OSD projects of Toshiba Solutions Corporation, since 2005 when its first prototype was developed. Hearing the voice of users, we have evaluated its application to ODS, and we have been improving the system and its rules and dictionaries. The evaluation of applying the system to OSD projects has been brought for the both side of OSD projects; the consignor in Japan and the consignee in China. The evaluation was carried out by analyzing the output of the proofreading system for documents and answers of questionnaires to the users who belong to system and software development division, and OSD management division in the consignor, and also to translators of the consignee. They appraise result of applying the proofreading system, by comparing original documents with their revised ones according to revision comments from the system.

Figure 4 shows the result ratio of comments from

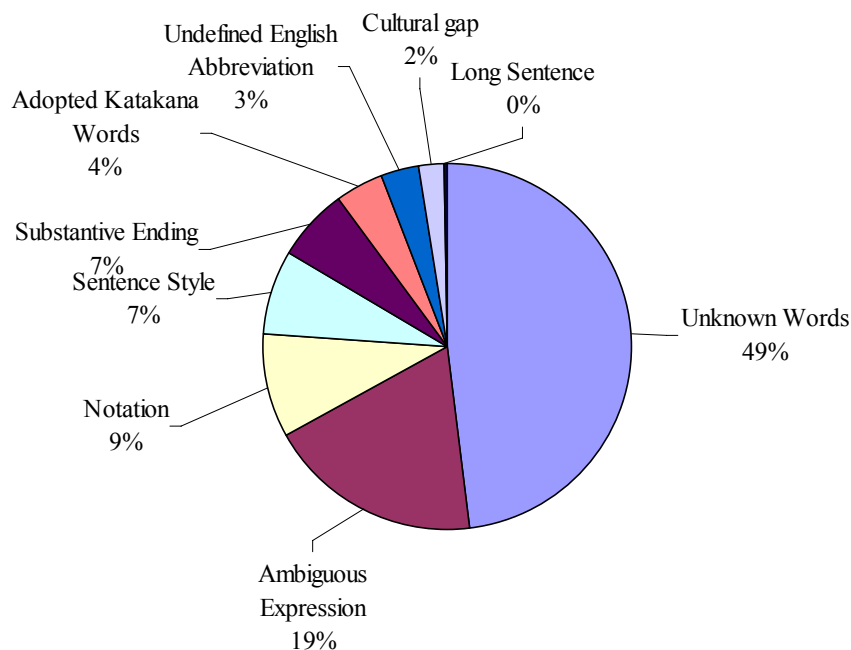


Figure 4. Comments provided by the proofreading system.

the proofreading system for actual specification documents for OSD written by Japanese engineers. 26,421 comments are provided for 168 documents.

Unknown words account for about half of the total number of comments, including misspelled words. Documents on OSD fundamentally write matters and items about latest computer and information technology, which are described using unfamiliar words. The documents describe specification of the systems applied in some specific application domain which contains a lot of particular terms of the domain. The dictionary of the system however is found insufficient. The system can however not distinguish misspelled terms from really unknown terms so far.

The following effects are found by applying proofreading of the OSD documents written by Japanese engineers.

(1) Comments on notations, sentence styles, and substantive endings provide basic quality of descriptions by authors.

The revised versions assisted with the comments have few unsuitable expressions, and concentrate attentions of authors and reviewers to implications of the documents. Reviewers were actually confronted with abundant usual mistakes, and problems of the content were buried behind them. The system is helpful to take those usual mistakes off, and to reveal crucial issues in the content.

(2) Expressions of the documents improve largely with less ambiguity.

Ambiguous expressions decrease in the documents and plain and definite expressions increase instead. The documents become easier to understand for non-natives. Japanese engineers after all recognize that

they write ambiguous expressions more than they thought, in many of their documents not only for OSD but for outsourcing to Japanese partners.

- (3) Comments concerning cultural problems are helpful for Japanese.

Cultural differences are difficult for ordinary Japanese (engineers) to recognize clearly. They often notice that some terms and things, which are very popular for Japanese, are particular in Japan only.

- (4) After delivery of the documents revised along the comments from the proofreading system, questions from Chinese engineers consequently decrease much.

Japanese engineers consumed much time to answer those many questions after they sent the specifications of the ordered software. Frequent questions and answers disturb their work, and the efficiency became worse.

The system is planned to be applied in the consignee company in China. The following effects are expected when the proofreading are applied for the OSD documents written by Chinese engineers.

- (1) The Chinese company expects improvement of basic quality of documents in Japanese.
- (2) Load of proofreading would be reduced for translators.
- (3) The engineers are expected to progress their Japanese language capability.
- (4) Managers of consignor side anticipate diminishing trivial mistake bothering them to review content.

Some problems are found to improve for the proofreading system;

- (1) Numbers of Comments provided by the system are too many.

The proofreading system provides comments on all terms and expressions which are possibly unsuitable, averse to taking risk. Users soon become accustomed to frequent comments, and inattentive to them. The evaluation method of expression suitability should be improved to suppress unnecessary comments.

- (2) Misspelled terms need to be distinguished from unknown terms.

A large number of unknown terms are detected by the system. Katakana terms, for example, are superfluously warned as unknown words. As mentioned above, those documents imply a number of technology terms originally in English. Those terms are written in katakana. Terms commented as unknown words include misspelled terms, which are difficult to discriminate from actual unknown words. The dictionary is insufficient for latest technology terms and specialized terms for application domain of the target software systems. The algorithm to detect misspell notation is to be improved.

Although katakana notations are usually very difficult for foreigners to understand, some technology terms

even in katakana are common also to Chinese software engineers. Too many comments on katakana terms disturb review by contrary.

- (3) Maintenance of the dictionary is not easy.

Supporting environment is necessary for the users to optimize the result of the proofreading system.

6. CONCLUSION

OSD (offshore software development) is typical international business collaboration. OSD projects often encounter intercultural and inter-linguistic problems disturbing the projects. Actual OSD projects contain various risks and problems for their success. One of the major factors of risks and problems is communication between the both shores. Business documents are formal media of information and knowledge for business collaboration. While documents for OSD should convey common understanding of the OSD project products, they often contain unsuitable and miss-understandable expressions and may draw risks, problems and offensive issues for the project. We propose proofreading system for OSD documents, focusing on OSD from Japan to China.

Japanese is a formal language for documents of OSD from Japan to China. In order to support business collaboration in offshore software development, we propose a proofreading system of documents written in Japanese to reduce unsuitable or inappropriate expressions. The system is used by Japanese outsourcers to refine documents written by Japanese engineers for Chinese readers, and also to be use by Chinese consignees and Japanese consignors to refine documents written by Chinese Engineers as derivatives of the projects.

Applying the system to actual OSD projects, the system is found very effective for reducing unsuitable expressions contained in the documents. Quality of documents was much improved. Q&As are consequently diminished. It improves efficiency of the OSD project for both shores. Engineers of Japanese outsourcer recognize ambiguity implied in their "specifications", not only for foreign partners but also for Japanese partners.

Support by applying proofreading system so far is limited to provide comments on unsuitable expressions. Users desire automatic rewriting or at least recommendation of rewritings to better expressions suitable and easy to read. We are investigating to realize it in order to support and to encourage international business collaboration.

ACKNOWLEDGMENT

The authors thank Mr. Seiji Iwata, Manager of the Business Intelligence Laboratory, and the lab's members, in the Advanced IT Laboratory, Toshiba Solutions Cor-

poration, for supporting this work.

REFERENCES

- Christiansen, M. (2007), Meeting the challenge of communication in offshore software development, *Software Engineering Approaches for Offshore and Outsourced Development First International Conference, SEAFOOD 2007, Zurich, Switzerland, Lecture Notes in Computer Science, Springer Berlin Heidelberg*, **4716/2007**, 19-26.
- Friedman, T. (2006), *The World is Flat (Updated and expanded): A Brief History of the Twenty-first Century*, Farrar Straus and Giroux (T).
- Iwata, S. (2005), Business document checking system for compliance, *Toshiba Review*, **60**(12), 36-39.
- Izuha, T. and Kumano, A. (2007), Chinese-to-Japanese Japanese-to-Chinese Machine Translation, *Toshiba Review*, **62**(4), 30-33.
- Kojima, S. and Kojima, M. (2007), Making IT offshoring work for the Japanese industries, *Software Engineering Approaches for Offshore and Outsourced Development First International Conference, SEAFOOD 2007, Zurich, Switzerland, Lecture Notes in Computer Science, Springer Berlin Heidelberg*, **4716/2007**, 67-82.
- Makino, K. (2005), Text mining technology for medical care field, *Toshiba Review*, **60**(9), 46-47.
- MIC (Ministry of Internal Affairs and Communications), Information and Communication Policy Bureau (2007), *Survey research report on progress of offshoring and its effects*, http://www.johotsusintokei.soumu.go.jp/linkdata/other017_200707_hokoku.pdf.
- Saito, K. (2007), Communication management for offshore development in China: *Key success factors of offshore development*, *Journal of the Society of Project Management*, **9**(1), 26-31.
- Sheng, Z., Nakano, M., Kubo, S., and Tsuji, H. (2008a), Experimental risk estimation for offshore software outsourcing, *IEEE Transactions on Electrical and Electronic Engineering*, **3**(3), 338-344.
- Sheng, Z., Tsuji, H., Sakurai, A., Yoshida, K., and Nakatani, T. (2008b), Preliminary analysis for risk finding in offshore software outsourcing from vendor's viewpoint, *Software Engineering Approaches for Offshore and Outsourced Development Second International Conference, SEAFOOD 2008, Zurich, Switzerland, Lecture Notes in Business Information Processing, Springer Berlin Heidelberg*, **16**, 134-148.
- Taniguchi, H., Zu, G., and K., Kano (2008), Contextual checking system for numerical inconsistencies in business documents, *Toshiba Review*, 2008, **63**(2), 70-73.
- Tsuji, H., Moriyasu, T., and Sheng, Z. (2008a), Offshore software development: its evolution stages and engineer's knowledge, *Journal of Information Processing Society of Japan*, **49**(5), 551-557.
- Tsuji, H., Nonomura, T., and Mibe, R. (2008b), Systems engineering approach for offshore software outsourcing, *Systems, Control and Information, Institute of Systems*, **52**(2), 54-59.
- Tsuji, H., Sakurai, A., Yoshida, K., Tiwana, A., and Bush, A. (2007a), Risk factors on offshore software development by conjoint analysis, *Transactions of Information Processing Society of Japan*, **48**(2), 823-831.
- Tsuji, H., Sakurai, A., Yoshida, K., Tiwana, A., and Bush, A. (2007b), Questionnaire-based risk assessment scheme for Japanese offshore software outsourcing, *Software Engineering Approaches for Offshore and Outsourced Development First International Conference, SEAFOOD 2007, Zurich, Switzerland, Lecture Notes in Computer Science, Springer Berlin Heidelberg*, **4716/2007**, 114-127.
- Wada, Y. and Tsuji, H. (2008), Critical factors analysis for offshore software development success by structural equation modeling, *Transactions of the Institute of Electrical Engineers of Japan. C, IEEE Transactions on Electronics, Information and Systems*, **128**(4), 540-545.
- Ye, Y., Nakakoji, K., and Yamamoto, Y. (2007), Reducing the cost of communication and coordination in distributed software development, *Software Engineering Approaches for Offshore and Outsourced Development First International Conference, SEAFOOD 2007, Zurich, Switzerland, Lecture Notes in Computer Science, Springer Berlin Heidelberg*, **4716/2007**, 152-169.
- Zu, G. (2007a), Specification checking system for Chinese offshoring, *Toshiba Review*, **62**(1), 70-71.
- Zu, G. and Kano, T. (2007b), The supporting technology of document proofreading to Japanese documents written by foreign engineers, *Proceedings of The 13th Natural Language Processing, 2007, The Association for Natural Language Processing, Tokyo, Japan*, 789-792.
- Zu, G., Taira, H., Makino, K., Kano, T., and Matsumoto, S. (2007c), The supporting technology of business document proofreading based on intercultural differences, *E-Commerce Technology and the 4th IE EE International Conference on Enterprise Computing, E-Commerce, and E-Services, 2007. CEC/EEE 2007. The 9th IEEE International Conference on*, 91-98.