

# A Japanese Corpus for Analyzing Customer Loyalty Information

Yiou Wang, Tahara Takuji

Fuji Xerox Co., Ltd

6-1, Minatomirai, Nishi-ku, Yokohama-shi, Kanagawa 220-8668, Japan

{Yiou.Wang, Tahara.Takuji} @fujixerox.co.jp

## Abstract

Today customers voice attitudes, opinions and their experience about some brands, companies, products or services through center calls, web reviews or SNS and analyzing them is an important task. On the other hand, customer loyalty has long been a topic of high interest in both academia and industry. Therefore, it is attractive to consider exploiting customer loyalty information by analyzing the voice of customer. However, although many previous studies focused on analyzing attitudes, opinions, sentiments of the text data, no work has been conducted from the perspective of customer loyalty, which is reflected by a combination of customer attitudes and behavior. In this work, we present JCLIC, Japanese Customer Loyalty Information Corpus, which is a corpus for analyzing customer loyalty information. For each review we have annotated detailed customer loyalty information which contains: *loyalty degree* that reflects loyalty level of the customer, *loyalty expression* that expresses the customer loyalty, *loyalty type* that indicates the category to which loyalty expression belongs., *reason expression* that expresses why the customer have such loyalty degree, and *reason type* that indicates the category to which reason expression belongs.. We describe our annotation scheme and annotation process, present results of an agreement study and give some statistics about the corpus we have annotated.

**Keywords:** Customer loyalty, Voice of customer, Opinion mining

## 1. Introduction

Customers voice attitudes, opinions and their experience about some brands, companies, products or services. Today voice of the customer (VOC), capturing a view of customer's behaviors, needs, and feedbacks, can be obtained through center calls, emails, questionnaire, web reviews or SNS (Subramaniam et al., 2012; Aguwa et al., 2012; Choi et al., 2013; Saeed et al., 2013). Such unstructured textual information contains valuable information for marketing analysis. On the other hand, customer loyalty has long been a topic of high interest in both academia and industry (Griffin, 2002; Asuncion et al., 2004; Scriosteanu and Popescu, 2010), and is critical for business profitability. Therefore, it is attractive to consider exploiting customer loyalty by analyzing the voice of customer.

To automatically find or track the attitudes, feelings and evaluations in texts, opinion mining and sentiment analysis have been extensively studied from different perspectives (Kushal et al., 2003; Pang and Lee, 2008; Nakagawa et al., 2008; Wang et al., 2012). There are also several corpus for opinion mining, such as MPQA Opinion Corpus version 3.0 (Deng and Wiebe, 2015), JDPA Corpus (Kessler et al., 2010) and Chinese evaluation information corpus (Wang et al., 2012). All the existing works are related to the affective and evaluative dimensions and focus on understanding customer satisfaction. However, customer loyalty is not customer satisfaction. Satisfaction is a necessary but not sufficient criterion for customer loyalty (Griffin, 2002; Ganiyu et al., 2012). We know that "very satisfied" to "satisfied" customers sometimes switch to competitors. Customer loyalty is multi-dimensional and comprise three dimensions namely, the emotive tendency towards the brand/product/service, the evaluative tendency towards the brand/product/service, the behavioral tendency towards the brand/product/service. To our knowledge, no work has been conducted from the perspective of customer loyalty, which is reflected by a combination of customer

attitudes and behaviors. There is also no corpora in this direction.

In this work we present a dedicated gold standard corpus of customer loyalty information. In order to investigate and understand how is customer loyalty and what drives loyalty in a particular market, we analyze the customer loyalty at a fine-grained level and define a 5-tuple that consists of (1) loyalty degree, (2) loyalty expression, (3) loyalty type, (4) reason expression, and (5) reason type as the basic unit of customer loyalty information. We annotate the texts of several domains (1500 reviews in 15 topics) in order to train a robust system and analyze a wide variety of customer loyalty information. To our knowledge, our corpus is the first resource of customer loyalty analysis in natural language field to date.

In the following, we will present our annotation scheme and annotation process. We will present results of an agreement study and give some statistics about the data we have annotated.

## 2. Customer Loyalty Information

**Customer Loyalty** is both an attitudinal and behavioral tendency to favor one brand over all others, whether due to satisfaction with the product or service, its convenience or performance, or simply familiarity and comfort with the brand (Griffin, 2002). Customer loyalty encourages consumers to shop more consistently, spend a greater share of wallet, and feel positive about a shopping experience, helping attract consumers to familiar brands in the face of a competitive environment.

The general purpose of this study is to identify, describe and analyze elements that have an impact on understanding customer loyalty or disloyalty. We need to investigate how is the customer loyalty degrees and types. The behavioral and attitudinal tendency of a customer is not consistent in some cases. Some customers have high evaluation and attitude but they are not repeat patronage, while some customers shop more consistently even though they are very

satisfied with the product. We need to distinguish such different loyalty types. Furthermore the factors which lead to loyalty or disloyalty need to be uncovered and understood before designing and implementing the strategies for customer retention and loyalty. Therefore we analyze the customer loyalty at a fine-grained level and define a 5-tuple that consists of (1) loyalty degree, (2) loyalty expression, (3) loyalty type, (4) reason expression, and (5) reason type as the basic unit of customer loyalty information. Each item is defined as follows and the details of the items will be described in Section 3.

- **Loyalty degree** is the level of loyalty or disloyalty.
- **Loyalty expression** is a span of text that describes the loyalty of a customer. It can be a single word, a multi-word expression, or a sentence.
- **Loyalty type** indicates the category to which loyalty expression belongs. We define the loyalty types into 7 categories and each type has a polarity information, positive (+), neutral (0) or negative (-). which will be described in Section 3.
- **Reason expression** is a span of text that describes the factors which lead to loyalty and disloyalty.
- **Reason type** indicates the category to which reason expression belongs. We define the reason types into 8 categories and each type has a polarity information, positive(+), neutral(0) or negative(-). which will be described in Section 3.

### 3. Annotation Schema

In order to obtain consistent annotations, we provided annotators with a more detailed annotation guidelines. Here we describe the annotation schema in general.

#### 3.1. Loyalty Degree

To understand customer loyalty one must recognize there are different degrees and types of loyalty. In order to investigate how is the customer loyalty degree, for each review, one loyalty degree is annotated.

Dick and Basu (1994) brought out the idea of relative attitudes while defining various forms of loyalty depicted below. They described loyalty as the strength of the relationship between a customer's relative attitude and repeat patronage and four dimensions had been identified: true loyalty, latent loyalty, spurious loyalty and no loyalty (Table1).

		Repeat Patronage	
		high	low
Relative Attitude	High	True Loyalty	Latent Loyalty
	low	Spurious Loyalty	No Loyalty

Table 1: Various forms of loyalty

Inspired by their idea, the loyalty degree is divided into 5 levels and is described in Table2. We annotate the loyalty degree by taking the loyal expression and loyalty types which will be described in Section 3.2 into consideration.

Level 2 is the same as true loyalty in Table1, which means to show loyalty in both behavioral and attitudinal dimensions. Level 1 include both spurious Loyalty and latent Loyalty in Table1. We extended the no loyalty and disloyalty into 3 levels in order to found the important differences between loyal and disloyal customers by text mining.

Level	Description
2	True or premium loyalty
1	Loyal customer or latent loyal customer
0	Neutral customer or No loyalty
-1	Latent churn customer
-2	Disloyal customer or churn customer

Table 2: The loyalty degree

#### 3.2. Loyalty Expression and Loyalty Types

Customer loyalty is the outcome of several factors comprising affective, behavioral and attitudinal dimensions. To clarify the scope of customer loyalty that we address in this study, we classify loyalty expressions into several categories. Such categorization is also helpful for further use of loyalty information.

Customer loyalty is reflected by a combination of attitudes and behaviors. The attitudes include how strong the psychological commitment or attachment is to the brand/service/product and emotional or feeling reactions in customer experiences. The behaviors include repeat purchasing, purchasing more and different products or services from the same company, and recommending the company to others. We define the loyalty types which include both behavioral and attitudinal aspects. In order to investigate the latent churn related information, the switching information are also annotated. The loyalty types are described in Table3.

Large Category	Loyalty Types
Attitude	Attachment
	Emotion
Behavior	Experience
	Repurchase
	Recommendation
Switching	Switching cost
	Competitor

Table 3: Types of loyalty expression

For each review, multiple loyalty expressions and the corresponding loyalty types should be annotated. Each type has a polarity information, positive (+), neutral (0) or negative (-). Positive is the loyal attitudes or behaviors, negative is the disloyal attitudes or behaviors. Here we give some examples to explain the loyalty expressions, the loyalty types and polarities. We use underline to show review topics and boldface for loyalty expression.

- Attachment +/0/-: an expression that expresses the condition of being attached to a product or a service.

- e.g. Product A is **my favorite one**. (Attachment+)
- Emotion +/0/-: an expression that expresses customer feelings or emotions.  
e.g. I **felt anger** at the clerk's rudeness in that shoe store. (Emotion-)
- Experience +/0/-: an expression that describes the experience of shopping or utilizing, such as how long or how often a customer has used a service .  
e.g. I **have been a member of this gym for several years** . (Experience+)
- Repurchase +/0/-: an expression that express repeat purchasing, purchasing more and different products from the same company.  
e.g. I will **buy other skin-care products** of this brand . (Repurchase+)
- Recommendation +/0/-: an expression that describes how customers rave about a product and promote it to their friends.  
e.g. I **don't recommend** this hotel at all. (Recommendation-)
- Switching cost +/0/-: an expression that shows the economic, psychological and physical costs to switch to the competitors.  
e.g. **It costs a lot to switch to** other mobile company. (Switching cost+)
- Competitor +/0/-: an expression that expresses the comparison with the competitor's product or temptation of the competitors.  
e.g. **The competitor's product is more convenient than** this healthcare wearable watch. (Competitor-)

Note that for some cases, our loyalty polarity may differ from the sentiment polarity. For examples, *The competitor's product is more convenient than this healthcare wearable watch*. Although the entire sentence is not negative, the polarity of loyalty expression, which implies the risk of churn, is negative. From this point of view, we consider the polarity in the connection with our topics and loyalty.

### 3.3. Reason Expression and Types

McKinsey & Company found(Griffin, 2002) :

- customers who have major problems but do not complain about them have a repurchase intention rate of about 9%
- those who do complain, regardless of the outcome, have a repurchase rate of about 19%
- customers who have their complain resolved have a repurchase intention rate of 54%
- customers who have complaints quickly resolved have a repurchase intention rate of 82%

Therefore in order to develop and sustain loyalty among the customers, it is important to find out what drives loyalty or disloyalty in a particular market. The factors which lead to loyalty need to be uncovered and understood. For each review, multiple reason expressions and the corresponding reason types should be annotated. Each type has a polarity

information, positive (+), neutral (0) or negative (-). Positive is the loyal factors, negative is the disloyal factors. The reason types is described in Table4.

Large Category	Reason Types
The product or service	Quality
	Function
	Price
Relating factors	Convenience
	Staff
	Related services
Other	External influences
	Others

Table 4: Reason types for reason expression

Here we give some examples to explain the reason expressions and the reason types. We use underline to show review topics and boldface for reason expression.

- Quality +/0/-: an expression that expresses the factors related to the quality.  
e.g. Camera X **broke just three days after I bought it**. (Quality-)
- Function +/0/-: an expression that is related to the function of a product or substantial service menu.  
e.g. Product A **can also perform a specific medical function**. (Function+)
- Price +/0/-: an expression that express price related factors.  
e.g. **It is too expensive** to go to the restaurant frequently. (Price-)
- Convenience +/0/-: an expression that expresses convenience, such as the location of the store  
e.g. The restaurant is **only five minutes away from my home**. (Convenience+)
- Staff +/0/-: an expression that describes the factors related to the attitude, skill level of staffs.  
e.g. The service desk staff in the gym **was very unfriendly** . (Staff-)
- Related services +/0/-: an expression that shows the reasons about aftercare, delivery and so on.  
e.g. **It took too much time for the delivery** since my order was placed in the web store. (Related services-)
- External influences +/0/-: an expression that expresses the influence by SNS or friends .  
e.g. I book this hotel **because of the review from Tripadvisor** . (External influences+)
- Others +/0/-: an expression that is loyalty factors which are not included in the above types.  
e.g. **Other users of the fitness club** were too annoying (Others-)

## 4. Corpus Construction

### 4.1. Data Collection

To train a robust system and analyze a wide variety of customer loyalty information, we constructed an customer loyalty information corpus which consists of Japanese reviews extracted from web pages of wide range of topics and

Mobile virtual network operator, Fitness club, Restaurant, English language school, Travel agents, Streaming services, Insurance services, Cloud services, Loan services, Beauty salon, Glasses, Wearable sensors, Laptops, Hospitals and dental clinics, Customer to Customer Websites, Airlines, Hair dryer, Watch, Hotels, Amazon Prime.

Table 5: Topics of Japanese loyalty information corpus

styles. We chose 20 topics which relate to products or services we use in daily life. The topics are shown in Table 5 (We only finished the annotation of the previous 15 topics). The followings are the steps for the corpus data collection: (1) Use the topic as the keyword and search documents using a Web search engine. (2) Collect HTML files of web pages from the retrieval results for each topic. Specifically, the pages in the retrieval results from forum sites, review sites are collected. We attempted to collect the reviews from different web pages. In this way, the corpus can cover different writing styles and reflect more diverse perspectives. (3) Randomly choose candidate reviews that are related to topic keywords from the above files. For each topic, we randomly collected 100 reviews, which may be in one sentences or in several sentences.

#### 4.2. Annotation Procedure

The followings are the steps for the construction of Japanese Customer Loyalty Information Corpus:

- Step1: Trained annotators checked whether the review contained any loyalty expressions or not. If the review contained loyalty expressions, the annotator annotated the text spans of the loyalty expressions, the types and the polarities of the loyalty expressions.
- Step2: The annotators checked whether the review contained any reason expressions or not. If the review contained reason expressions, the annotator annotated the text spans of the reason expressions, the types and the polarities of the reason expressions.
- Step3: Trained annotators judged the loyalty level of a whole review and assigned a level label from a predefined set.

In some cases, one review may contain multiple loyalty expressions or multiple reason expressions, and then multiple loyalty information sets must be annotated. That is to say, all the customer loyalty information was annotated for each review. It took 360 man-hours to construct the corpus in 15 topics.

Table6. shows one example of annotation result of the following review.

**The original review:** *My daughter is going to this English school twice a week. After she studied in the school for half a year, her Katakana English became to the natural accent. The teachers always find a way to make the class funny and interesting. Going to the English school is very*

*enjoyable for her. I think this school is the first choice for all beginners.*

Loyalty expression	Loyalty type
is going to this English school twice a week	Experience+
Going to the English school is very enjoyable	Emotion+
this school is the first choice for all beginners.	Recommendation+
Reason expression	Reason type
her Katakana English became to the natural accent	Quality+
The teachers always find a way to make the class funny and interesting.	Staff+

Loyalty degree

2

Table 6: One example of annotation result

## 5. Corpus Analysis

### 5.1. Statistics

We have finished the annotation of 1500 reviews in 15 topics, including 7,780 sentence. Statistics about our complete corpus can be found in Table7.

Topics	15
Review	1,500
Loyalty reviews	831
Disloyalty reviews	570
Neutral reviews	99
Sentence	7,780
Loyalty expression	3,922
Reason expression	4,882

Table 7: Statistics of the corpus

As for some detailed statistics, the information of loyal expressions and types is shown in Table8 and the information of reason expressions and types is shown in Table9. The result shows that emotion type is the most in loyalty expressions. The number of loyalty expressions which show attitudinal tendency is 2,232, while the number of loyalty expressions which show behavioral tendency is 1,222. The remaining 468 loyalty expressions are about switching information. The loyalty expressions reflected by customers' attitudes covers the majority. As for the reason expressions, the reasons related to "staff" are the main factor and the following is the quality.

### 5.2. Inter-Annotator Agreement

We first let both annotators independently annotate the same 2 reviews in English school and beauty salon topics, on which we compute the inter-annotator agreement. After that, each of the annotators annotated their own set of 13 reviews. We take one of the annotator's annotation as a

Loyalty Types	Numbers
Attachment	406
Emotion	1,826
Experience	693
Repurchase	308
Recommendation	221
Switching cost	43
Competitor	425

Table 8: Statistics of loyalty expressions and types

Reason Types	Numbers
Quality	1,141
Function	936
Price	504
Convenience	294
Staff	1,326
Related services	340
External influences	120
Others	221

Table 9: Statistics of reason expressions and types

gold standard and the other annotator’s annotation as a system output, and then use the following measures to evaluate inter-annotator agreement:

- **Recall:** ratio of correctly extracted loyalty/reason expressions to the number of expressions in the gold standard.
- **Precision:** ratio of correctly extracted loyalty/reason expressions to the number of expressions in system output.
- **F-measure:** harmonic mean of recall and precision.
- **Accuracy:** ratio of the number of correct system output to the number in the gold standard.

For the annotation of text spans where each annotator individually picks some words from the review instead of assigning a label from a predefined set, we follow Breck et al. (2007) and use the following partial match criteria to calculate the recalls, precisions and F-measures of the loyalty/reason expressions. The partial match is defined as follows:

- **Partial match:** extracted expression is regarded as correct if it overlaps the gold standard’s one.

The inter-annotator agreement is shown in Table10. With the only exception of reason type accuracy of beauty salon, for the annotation of assigning a label from a predefined set, such as loyalty degree, loyalty type and reason type, we have achieved a reliable accuracy, ranging from 80% to 90%. For the annotation of extract some words from the review, the F-measures are ranging from 75% to 80%. The detailed results from the agreement study will be used to refine the annotation guidelines.

	English School	Beauty Salon
Loyalty Degree Accuracy	0.85	0.96
Loyalty Expression Recall	0.66	0.67
Loyalty Expression Precision	0.85	0.86
Loyalty Expression F-measure	0.74	0.75
Loyalty Type Accuracy	0.85	0.85
Reason Expression Recall	0.63	0.93
Reason Expression Precision	0.90	0.70
Reason Expression F-measure	0.74	0.80
Reason Type Accuracy	0.83	0.57

Table 10: Inter-annotator agreement

## 6. Conclusion

In this work, we presented a dedicated gold standard corpus of customer loyalty information in Japanese. For each review we have annotated detailed information about the customer loyalty. We have described our annotation process and given an overview of our annotation guidelines. In total, we have annotated 7780 sentences from reviews in 15 topics. We will extend our corpus and build supervised models for determining the loyalty level, recognizing loyalty expressions and reason expressions and classifying the loyalty types and reason types in further.

## 7. Bibliographical References

- Aguwa, C. C., Monplaisir, L., and Turgut, O. (2012). Voice of the customer: Customer satisfaction ratio based analysis. *Expert Systems with Applications*, 39:10112–10119.
- Asuncion, B., D.J.Martin, and Quintana, A. (2004). Model of customer loyalty in the retail banking market. *European Journal of Marketing*, 38:253–275.
- Breck, E., Choi, Y., and Cardie, C. (2007). Identifying expressions of opinion in context. In *Proceedings-IJCAI-2007*, pages 2683–2688.
- Choi, C., J. Lee, G. P., Na, J., and Wan-SupCho. (2013). Voice of customer analysis for internet shopping malls. *International Journal of Smart Home*, 7:291–304.
- Deng, L. and Wiebe, J. (2015). Mpqa 3.0: Entity/event-level sentiment corpus. In *2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT’15)*, Colorado, USA.
- Dick, A. S. and Basu, K. (1994). Customer loyalty: Toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22:99–113.
- Ganiyu, R. A., Uche, I. I., and Elizabeth, A. O. (2012). Is customer satisfaction an indicator of customer loyalty? *Australian Journal of Business and Management Research*, 2(7):14–20.
- Griffin, J. (2002). *Customer Loyalty: How to Earn It, How to Keep It*. Jossey-Bass, 2nd edition.
- Kessler, J. S., Eckert, M., Clark, L., and Nicolov, N. (2010). The 2010 ICWSM JCPA sentiment corpus for the automotive domain. In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC’10)*.
- Kushal, D., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and se-

- semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*.
- Nakagawa, T., Kawada, T., Inui, K., and Kurohashi, S. (2008). Extracting subjective and objective evaluative expressions from the web. In *Proceedings of the 2nd International Symposium on Universal Communication*, pages 251–258.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Saeed, R., Lodhi, R. N., Munir, J., Riaz, S., Dustgeer, F., and Sami, A. (2013). The impact of voice of customer on new product development. *World Applied Sciences Journal*, 24:1255–1260.
- Scriosteanu, A. and Popescu, D. (2010). Customer loyalty - the major goal of customer relationship management. *Annals of University of Craiova - Economic Sciences Series*, 3:294–300.
- Subramaniam, L. V., Faruquie, T. A., Iqbal, S., Godbole, S., and Mohania, M. K. (2012). Business intelligence from voice of customer. In *Proceedings of IEEE International Conference on Data Engineering (ICDE'09)*, Shanghai, China, March. IEEE.
- Wang, Y., Junichi, Kazama, Kawada, T., and Torisawa, K. (2012). Chinese evaluative information analysis. In *Proceedings of the 24rd International Conference on Computational Linguistics (COLING'12)*, pages 2773–2788.