Asia Pacific Journal of Information Systems
Vol. 31 No. 4 (December 2021), 527-542

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2021.31.4.527

The Impact of Insurance Contract on Insurance Complaint Ratios through Text Analysis

Jeongkwon Seo[a], Woojin Yang[b,*], Hyejin Mun[c], Chul Ho Lee[d]

[a] Master Student, College of Business, Korea Advanced Institute of Science and Technology, Korea
[b] Ph.D. Student, College of Business, Korea Advanced Institute of Science and Technology, Korea
[c] Ph.D. Student, College of Business, Korea Advanced Institute of Science and Technology, Korea
[d] Assistant Professor, College of Business, Korea Advanced Institute of Science and Technology, Korea

ABSTRACT

The government-driven open data policies are on the rise to protect consumers from misunderstandings and monitor the companies. However, in contract-based industries such as insurance, the contract-inherent characteristics make information asymmetry between consumers and companies. Our paper focuses on insurance contracts where the contingency has high uncertainty of occurrence, and the clauses may incur high costs of reading. Given those contracts, we hypothesized that the contract's clear statement decreases customer dissatisfaction and lowers the number of complaints. To empirically support the claim, we collected customers' complaint documents of insurance companies and insurance contracts from 2005 until 2017. Our econometric models showed that clearer statements and words significantly reduce the complaints after controlling for firm-specific heterogeneity and time-specific heterogeneity. We identify that insurance companies' complaint ratio significantly differ depending on the insurance contract, including specific clauses and words.

Keywords: Incomplete Contract Theory, Consumer satisfaction, Ambiguity, Insurance Contract

Ⅰ. Introduction

The government's efforts for digitalization worldwide are accelerating than ever before to realize public interest and increase citizen trust (World Bank Blogs, 2021). In particular, due to the COVID-19 crisis, the demand for launching and activating digital government and the use of digital technology are increasing. Moreover, the efforts of digital government on data sharing have social and economic purposes. The social purpose is meant to achieve through transparency enhancement and anti-corruption, and the economic purpose is to generate profits and increase efficiency from the public sector to the industrial sector. Likewise, the trend of digitalization affects individual industries as well as the govern-

*Corresponding Author. E-mail: dnwls2669@kaist.ac.kr ment's direction. As stated in the McKinsey report (2017), the insurance industry is transitioning through the movement into digitalization. Digitalization within the industry expands the size of the industry and is essential for new initiatives to increase profits. In addition, digitalization enables the insurance industry to prepare for the future direction by identifying strategic strengths and potential threats through data sharing and utilization.

In line with digitalization, the department of insurance in the U.S. makes public information about insurance firms to relieve tension between consumers and companies[1]. They strive to protect consumers from the lack of information and monitor the companies. The responsibilities include comprehensive financial oversight, products review, policy forms, company practices for compliance with the laws, and investigation of consumer complaints about insurance companies (Andrew, N. M., 2019). Despite the efforts by the governments, the characteristics of the insurance contract itself make to keep a distance between consumers and insurance companies.

Contracts are inherently unable to control all market players' activities due to unpredictable and cost-prohibitive situations. Incomplete contract theory (Coase, 1937) eases contract completeness and recognizes that a contract causes information asymmetry between providers and consumers. Information asymmetry leads the two parties to interpret the same signal differently (Hart, 2017). The concern on information asymmetry emerges over contract incompleteness in the insurance industry. According to the 2019 Kiri report, the rate of insurance registration per household is about 98.2 %, especially when registering car insurance is 75.5 %. Despite the high registration rate, a new Deloitte survey found that over 90 % of consumers agree with legal terms and obligations without the careful process of reading them.

Insurance contract comprises rights and obligations between parties, punishment for violations, and duration, and the contract is based on the relationship between service providers and customers. The most common reasons for not reading the terms carefully are the large number of pages, the uniqueness of terminology, complexity, and difficulty of terms and sentences. Moreover, texts within contract bear ambiguity because there is a possibility that an individual word or sentence can be interpreted in various meanings (Abraham, 1996; Barry and Thomas, 2018; Boardman, 2005; Knutsen, 2010). Burton (2018) argued that formal texts such as contracts should be written in an easy language where messages can be delivered without conflicts occurring from lexical ambiguity.

Most previous studies in the insurance industry context were conducted with respect to the determinants of consumer satisfaction and the characteristics of insurance types. Doerpinghaus (1991) identified that companies guaranteeing more support for drivers with higher risks have a greater dissatisfaction rate with empirical evidence across two states, Oregon and Illinois. According to James et al. (2005), companies that retain more legal costs for disputes with customers and a comparably higher budget for car insurance have fewer complaints.

A complaint is defined as written correspondence expressing a grievance against an insurer or carrier, and claim handling is the essential part of auto insurance complaints. Claim handling contains unsatisfactory settlement offers, which are denials of claims related to the insurance contract. For example, if the contract does not include the coverage a customer

[1] The National Association of Insurance Commissioners (NAIC), Our Focus: Our Members and our Mission

filed a claim under, the claim is denied without insurance plan compensation. Likewise, complaints could arise from ambiguity of text which links to the unsatisfactory settlements of compensation from the different interpretations.

Despite a need to understand the role of legal contracts on consumer satisfaction, no study has highlighted the relationship between contracts and complaints. Our study starts with the following questions: 1) Does the insurance policy affect consumers' complaints? 2) What features of the contract affect consumer dissatisfaction? Therefore, this study investigates whether the insurance contract is associated with consumers' complaints and what kinds of linguistic features within the contracts affect insurance complaints.

To empirically examine the association, we collected text data of auto insurance policies, complaint ratios against each car insurance firm, and the considered insurance firms' heterogeneous characteristics from the U.S. Department of Insurance (the number of coverages, premium, and market share). Using the firm-state-year panel data, we identified firms' unobserved heterogeneous effects exist, although observed characteristics are controlled.

Moreover, to estimate the contracts' impact on complaint ratios, we assessed the policies from three perspectives: overall ambiguity, word class, individual words. Starting with the degree of ambiguity in a given contract, we gradually narrowed the scope of independent variables by word-class and individual words. The result suggested that the complaint ratios increase as the degree of ambiguity increases; on the other hand, the complaint ratios decrease with the use of the pronoun, which would clarify the subjects. Furthermore, the words related to a condition (exclude, include, must, if) also reduce complaint ratios as the role of clarifying the requirements of compensation.

The rest of this paper is organized as follows. Section 2 discusses previous literature and theoretical framework, and Section 3 explains Research Methodology. In Section 4, the work related to the data analysis and econometric analysis is described. Furthermore, we provide an overview of the results and discussion. Finally, the limitations and implications of this study appear in Section 5.

Ⅱ. Conceptual Background

2.1. Incomplete Contract Theory

Contracts are not able to be complete even if they pursue the form of a complete contract. Everyone does not accept a contract as a complete meaning, and the languages used in the contract could be ambiguous or unclear. According to Hart (2017), property rights theory explains that property rights can lead to an optimal allocation of the contract. Moreover, property rights determine the size of the gains from a contract. Property rights theory notes if the asset ownership skewed on one side, the hold-up problem would occur. Hart (2017) investigated the disadvantageous side inevitably renegotiate or be punished with the court if the Hold-up problem occurs. Due to the Hold-up problem, the deadweight loss also inevitably happens in the renegotiation process. There are various pieces of research on moral hazard, adverse selection, and insurance fraud problems caused by the customer. Furthermore, so many examples exist with articles and research for hiding or ambiguous insurance contracts leading to unsatisfying compensation (Kim, 2012).

Hart's property rights theory can explain this situation. First, the insurance company has property rights, then they do entitlement or shading the information. Second, the customer cannot fully understand the insurance contract, so their expectation surplus is relatively large compare to actual compensation. Moreover, premium (price) will be set by this expectation surplus and expected compensation. When an accident happens, the actual customer surplus will be down, and the insurance company's cost also down. The actual customer surplus will be lower than the premium, and the insurance company's surplus will be larger than contract timing. So, a hold-up problem has occurred. Customers can renegotiate with the insurance company or lawsuit to court. However, renegotiating is not practical because property rights are so strong, and the customer has no information about the complex, ambiguous contract. Then customers receive the compensation or take the lawsuit to court. So, the information asymmetry between company and customer increases as the complaint ratio related to insurance contracts increases.

Likewise, contracts are created for solving the moral hazard problems that come from information asymmetry. However, these contracts also bring another moral hazard problem due to their incompleteness. Thus, due to the limitations of contracts, cost-ineffectiveness, the start of incomplete contract theory is justified from the research of Coase (1937). Regarding assigning control power and rights of the insurance contract that is naturally complex and contains legal information, the insurance company side may be advantageous because the party contributing more to the value created by contracting gets control rights.

2.2. Complaints in Insurance Industry

A complaint is defined as written correspondence expressing a grievance against an insurer or carrier, and it results in a written request for information from the company. Claim handling is the essential part of auto insurance complaints, and claim handling contains unsatisfactory settlement offers, which are denials of claims related to the insurance contract. For example, if your contract does not include the coverage you filed a claim under, your claim is denied without getting any compensation from your plan. The complaint ratio is used as a proxy for measuring consumer satisfaction in other literature (Karl and Wells, 2016; Pawell, 2017).

Moreover, previous studies have been investigated a variety of factors influencing the complaint rate of insurance customers. Doerpinghaus (1991) showed that consumers are more likely to have dissatisfaction with insurance companies that guarantee more coverage for high-risk drivers. Also, Carson (2005) has proven that companies with more legal costs for disputes with customers and have a higher budget portion for car insurance against other insurance have fewer complaints. Kim and Yang (2019) analyzed that the factors affecting consumer complaints and dissatisfaction in auto insurance are lack of compensation, lack of understanding of contract details, and late notice for contract changes.

2.3. Insurance Contract and Ambiguity

Linguists define ambiguity as expressions that could be interpreted in multiple ways. Given the nature of natural languages, linguists have commonly assumed that natural languages are ambiguous (Garrett, 1970). In legal contracts, ambiguity issue is critical. Ambiguous expressions caused by the wrong position of words in a sentence and unclear use of modifiers may generate legal disputes. Ambiguity does not appear on the surface. That is why each party interprets the contract in their own interests after an unexpected event occurs (Torbert, 2014).

Insurance contracts and conditions refer to documents that insurance companies prepared in advance to contract with consumers. These documents' contents include the terms, rights, and obligations between two parties having a contract. It also consists of sanctions imposed when one party does not carry out the obligations required. However, due to the nature of terms and conditions written in letters, the ambiguity of words and phrases brings difficulty to have a consistent interpretation for both parties when not reaching an agreement.

In particular, misinterpretation of insurance policies may even lead to legal disputes because the terms and conditions in the insurance industry include the subject party, the duration, and scope of compensation guaranteed, the coverage, and the restriction. For example, in respect to collateral damage protection, the contract for auto insurance covers that the insurance company compensates the insured for direct damage to the insured vehicle from the following accidents during the possession of the insured vehicle.

In such cases, accessories and machinery parts generally attached to or equipped with an insured motor vehicle are considered part of the insured motor vehicle. However, what is not usually attached or equipped is limited to what is stated in the insurance contract. A recent dispute has been made over the damage to the panorama sunroof fitted to the motor vehicle. Likewise, an insurance contract inevitably includes some ambiguous phrases and sentences.

2.4. Ambiguity Type

Ambiguity is classified depending on the sources. Massey et al. (2014) outline six distinct types of ambiguity: Lexical ambiguity, Structural ambiguity, Semantic ambiguity, Vagueness, Incompleteness, and Referential ambiguity. Lexical ambiguity is the most common ambiguity, which occurs when a word or phrase has two or more valid meanings. Second, structural ambiguity occurs when sentence structures allow sentences to be interpreted in multiple ways. It is generated due to the association and placement among words and clauses in a sentence. Semantic ambiguity occurs when a sentence has multiple interpretations in its context (i.e., Kim and Lee are married). Vagueness refers to ambiguity when a statement allows borderline cases or relative interpretation. Incompleteness occurs when a statement fails to convey a single clear interpretation because of too little detail. Referential ambiguity occurs when a word or term in a sentence does not have an apparent reference.

We focused only on lexical and structural ambiguity since both are identifiable with a single word's number of meanings and its word class. In this paper, we limit text analysis to a single word, not a phase or a sentence.

2.5. POS tagging & Text mining

In corpus linguistics, part-of-speech tagging, also called grammatical tagging, is the process of marking up a word in the text as a particular part of speech, based on both its definition and its context of words as nouns verbs, adjectives, adverbs, etc. NLTK POS tagging library classifies these words in sentences to POS tag (Bird, 2006). In addition, NLTK provides libraries for natural language processing. Prior literature has used POS tagging for analyzing and extracting the information or features in sentences or documents. Opinionated words in the blog text are extracted by POS tagging (Singh et al., 2011). This paper also employs POS tagging for setting the word class.

Moreover, text mining tools are also usually used to analyze formal documents. For example, they evaluate the security risk factors by extracting information related to disclosure which is positive or negative (Wang et al., 2013). Macro (2020) extracts information from the internal banking contract document with the NLTK POS tagging library. They use the default NLTK POS tag that is based on the Penn Treebank corpus (Marcus, 1993).

Ⅲ. Research Methodology

3.1. Data Source and Data Sampling

To empirically assess the impact of an insurance contract, we collected the three major data: text format of auto insurance policies, reported complaints of focal insurance firms, and the heterogeneous characteristics of the firms. The primary source is the U.S. Department of Insurance. The U.S. government releases the information of insurance companies to protect consumers from misunderstandings, assist in the consumer's purchasing decision process, and monitor the companies under state laws. Likewise, the government is responsible for easing the information asymmetry between the company and consumers. In this context, government-driven open data is appropriate for investigating the consumers' complaints in that its purpose matches our interests.

We obtained cross-sectional information of insurance policies and the number of reported complaints against focal firms given a year and state. We then standardized the complaints by dividing them into the premium of the focal firm resulting complaint ratio. Previous studies on insurance complaints have mainly used this measure (James et al., 2005). The observable characteristics of insurance firms include market share and the number of coverages. We then constructed the firm-state-year panel data set. Text information related to policies' features is cross-sectional, which means it is fixed across time and states. There was not much change in insurance policies during the sample period. Sample for analysis subjected to 21 automobile insurance firms spanning from 2005 to 2017 across 8 states. We targeted automobile insurance because the related issues with respect to compensation coverage are frequently raised in damage insurance. Further, automobile insurance accounts for the most considerable portion of damage insurance[2].

3.2. Empirical Methodology

In the baseline analysis, we identify whether firms' unobserved heterogeneous effects on complaint ratio exist, although observed characteristics are controlled. We regard these unobserved effects as the effects of insurance policies. The following equation indicates the baseline model:

$$ComplainRatio_{i,j,t} = \beta_{0} + \sum \beta Company_i + S_j + Y_t + \gamma X_{i,j,t} + \epsilon_{i,j,t} \qquad (1)$$

The dependent variable is complaint ratio against firm *i* located in state *j* at year *t*. $Company_i$ is dummy variable representing each insurance firms. State and year fixed effects are included in the model as S_j and Y_t, respectively. $X_{i,j,t}$ represents the firm's characteristics, including the number of coverages that

[2] According to Statista Research Department, State Farm Automobile Insurance was ranked first with a 9.3 percent market share in 2019.

the firm serves and market share. β informs heterogeneity in complaint ratio across 21 insurance firms after considering observable variables. This heterogeneity is attributable to different policies across firms.

Further, in order to estimate the impact of policies on complaint ratios, we assess the policies from three perspectives: Lexical ambiguity, structural ambiguity, and individual words. Starting with the ambiguity of a given overall contract for assessing the effect of lexical ambiguity, we gradually narrow the scope of independent variables by word class for structural ambiguity and individual words.

To measure the degree of lexical ambiguity in a contract, we employed the ambiguity index, which was calculated by the sum of each word multiplying the word frequency in the contract and the number of focal words' meaning then divided by total words frequency. A researcher assessed the lexical ambiguity using the number of meaning counts (Britton, 1978; Ceccato, 2004). The ambiguity index implies how ambiguous an insurance contract is, which could confuse customers. The following equation is to estimate the impact of lexical ambiguity on complaint ratio:

$$ComplainRatio_{i,j,t} = \beta_0 + \beta_1 Ambigulity_i + S_j + Y_t + \gamma X_{i,j,t} + \epsilon_{i,j,t} \qquad (2)$$

$Ambiguity_i$ indicates ambiguity index of firm i's insurance contract, which measures how ambiguous an overall contract is to customers. β_1 is our interests and we hypothesize that the degree of ambiguity raises up complaints of customer resulting positive effect.

More specifically, we extracted word classes from each contract by using POS tagging to assess structural ambiguity. POS tagging in corpus linguistics, part-of-speech tagging, also called grammatical tagging, marks up a word in the text as a particular part of speech, based on its definition and context: nouns, verbs, adjectives, adverbs, etc. Using the library provided by NLTK, we firstly tokenized the insurance contract document. Using POS tagging, we were able to get the words' POS tag in the insurance policies. Then, we counted the POS frequencies in each company's insurance contract. Finally, we put these POS tags in Equation (3):

$$ComplainRatio_{i,j,t} = \beta_0 + \beta_1 WordClass\frac{k}{i} + S_j + Y_t + \gamma X_{i,j,t} + \epsilon_{i,j,t} \qquad (3)$$

$WordClass\frac{k}{i}$ represents the frequency of word-class k in the contract of firm *i*. Our study focused on the following word-class regarding structural perspective: Coordinating Conjunction (CC), Adjective (JJ), Possessive Ending (POS), Pronoun (PRP), Preposition (IN), and Cardinal Digit (CD). For example, Coordinating Conjunction such as "and" or "but" may cause structural ambiguity in that a sentence can be interpreted in more than one way depending on sentence structure.

Lastly, we estimate the impact of individual words directly related to compensation on complaint ratios. Insurance policies are required to clarify the coverage, condition, and limitation with respect to payment or compensation. It is plausible that more such words reduce complaint ratios against the focal firm. Thus, we counted words related to condition across policies and constructed the following OLS equation:

$$ComplainRatio_{i,j,t} = \beta_0 + \beta_1 Word\frac{m}{i} + S_j + Y_t + \gamma X_{i,j,t} + \epsilon_{i,j,t} \qquad (4)$$

$Word\frac{m}{i}$ indicates the frequency of individual words m in firm *i* contract. We assess the individual words that would be associated with compensation, such as “exclusion”, “if”, “damage”, “must”, etc.

Ⅳ. Data Analysis and result

4.1. Baseline Result

<Table 1> is the results of Equation (1) to estimate

<Table 1> Firm Heterogeneity in Complaint Ratio

Dependent Variable: Complaint ratio	(1)	(2)	(3)
	Model1	Model2	Model3
Market share			-0.653
Num. of coverage			0.245***
Firm-specific Heterogeneity			
Allied	-0.531***	-0.549***	-0.0570
Allstate	-0.646***	-0.563***	-0.997***
American	-0.415***	-0.358***	-0.0852
Amica	-0.640***	-0.616***	-0.367***
CSAA	-0.640***	-0.402**	-0.890***
Esurance	-0.001	0.0514	0.298**
Farm Bureau	-0.370	-0.376	0.144
Farmers	-0.637***	-0.564***	-0.774***
Geico	-0.482***	-0.426***	-0.886***
Hartford	-0.329**	-0.286**	0.211
Liberty	-0.406***	-0.348***	-1.314***
Loya	0.993***	1.073***	1.074***
Metropolitan	-0.536***	-0.465***	0.033
Progressive	-0.584***	-0.513***	-0.456***
Safeco	-0.620***	-0.535***	-0.512***
Shelter	-0.654***	-0.522***	-1.256***
State Farm	-0.701***	-0.618***	-0.988***
Travelers	-0.372**	-0.392***	-0.877***
Unique	2.805***	3.101***	3.347***
USAA	-0.543***	-0.506***	
Constant	0.920***	1.269***	-1.186*
State Fixed	No	Yes	Yes
Year Fixed	No	Yes	Yes
F-value	14.43	11.08	10.83
N	726	726	726
adj. R-squared	0.270	0.363	0.363

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns: insignificant at the 0.05 level the firm-heterogeneity. Column 1 shows OLS (Ordinary Least Squares) regression results without any control variables and fixed effects. The coefficients of each firm dummy variable yield significant value except for 'Esurance' and 'Farm Bureau', indicating the existence of firm-heterogeneity on a complaint ratio. Column 2 includes state and year fixed effects with Column 1, and the results are consistent with the former one. Also, the third column represents the full model of Equation (1). Market share does not significantly affect the complaint ratio, whereas the number of coverages positively affects the complaint ratio. According to previous studies, companies with more insurance coverages for high-risk drivers have higher complaint ratios (Doerpinghuas, 1991). It can also be interpreted that the complaint ratio becomes higher as the number of coverages increases because the company includes the above contents, or the company cannot guarantee many types of coverage compensation. Although observable characteristics were controlled, such as market share, number of coverages, year- and state-specific effect, the coefficients are significantly heterogeneous across insurance firms. We presume that these differences were due to a firm-specific insurance contract. Heterogeneity in complaints level justifies additional analysis of insurance policies formed as a contract in the diverse perspective.

4.2. Ambiguity Index

<Table 2> is the results on the impact of lexical ambiguity on the complaint ratio used by Equation (2). The results are consistent across all models that complaint ratio is positively associated with ambiguity index. As mentioned above, the ambiguity index indicates how lexically ambiguous an insurance contract is. The positive effect on consumers' complaints

<Table 2> Impact of Ambiguity on Complaint Ratio

Dependent Variable: Complaint ratio	(1)	(2)	(3)
	Model1	Model2	Model3
Ambiguity index	0.437*** (0.106)	0.404*** (0.102)	0.288** (0.118)
ln(Word count)			-0.040 (0.056)
Market share			-1.367** (0.562)
Num. of coverage			-0.008 (0.0156)
Constant	-3.000*** (0.843)	-2.545*** (0.843)	-1.115 (1.004)
State Fixed	No	Yes	Yes
Year Fixed	No	Yes	Yes
F-value	17.18	4.29	4.14
N	726	726	726
adj. R-squared	0.022	0.091	0.098

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns: insignificant at the 0.05 level means that difficulties or confusion in interpreting policies may lead to customer dissatisfaction. Asymmetric information induced by incomplete contracts can make customers disadvantaged and generate dissatisfaction by the customer.

4.3. Word Class

<Table 3> shows the results of word classes to assess structural ambiguity. Column 1 to Column 6 indicates the results of each word class: Coordinating Conjunction (CC), Adjective (JJ), Possessive Ending (POS), Pronoun (PRP), Preposition (IN), and Cardinal Digit (CD), respectively. Complaint ratio increases with the number of Coordinating Conjunction and Preposition classes. Both can make multiple meanings of a sentence depending on the displacement within a sentence, resulting in higher structural ambiguity (Berry et al., 2003; Hindle, 1993; Tobert, 2014). In contrast, the frequency of Pronoun

<Table 3> Impact of Word Classes on Complaint Ratio

Dependent Variable: Complaint ratio	(1)	(2)	(3)	(4)	(5)	(6)
	Model1	Model2	Model3	Model4	Model5	Model6
ln(CC)	0.563* (0.298)					
ln(JJ)		0.215 (0.243)				
ln(POS)			-0.012 (0.023)			
ln(PRP)				-1.016*** (0.106)		
ln(IN)					1.480*** (0.428)	
ln(CD)						-0.996*** (0.185)
ln(Word count)	-0.531* (0.284)	-0.202 (0.231)	0.014 (0.062)	1.098*** (0.126)	-1.513*** (0.440)	0.898*** (0.176)
Market share	-1.961*** (0.513)	-2.081*** (0.541)	-2.076*** (0.581)	-3.338*** (0.506)	-0.376 (0.598)	-1.509* (0.510)
Num. of coverages	-0.008 (0.016)	-0.012 (0.016)	-0.011 (0.016)	-0.019 (0.015)	-0.016 (0.015)	0.017 (0.016)
Constant	2.190** (0.901)	1.314* (0.755)	0.760 (0.633)	-2.855*** (0.675)	4.283*** (1.139)	-2.012*** (0.786)
State Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed	Yes	Yes	Yes	Yes	Yes	Yes
F (25, 700)	4.03	3.91	3.88	8.03	4.42	5.19
N	726	726	726	726	726	726
adj. R-squared	0.095	0.091	0.091	0.195	0.105	0.126

Note: *p < 0.05, **p < 0.01, ***p < 0.001, ns: insignificant at the 0.05 level and Cardinal Digit classes is negatively associated with the complaint ratio. Pronouns are used to specify subjects such as customers or company thereby reducing the structural ambiguity. In addition, Cardinal Digit may contain section numbering and exact numerical value, which can reduce the ambiguity. Likewise, the results show that overall, lexical and structural ambiguity are positively associated with the complaint ratio. It supports that asymmetric information induced by incomplete contract can make a customer be disadvantaged position and generate dissatisfaction by the customer.

4.4. Individual Words

We analyzed individual words which are classified

<Table 4> Impact of Individual Words on Complaint Ratio

Dependent Variable: Complaint ratio						
POS	word	coef. (std.)		POS	word	coef. (std.)
Coordinating Conjunction (CC)	but	0.01 (-0.08)		Preposition (IN)	under	0.40*** (-0.11)
	however	0.02 (-0.04)			within	0.55*** (-0.11)
	or	0.16 (-0.24)			as	0.93*** (-0.16)
	and	0.25* (-0.13)			of	1.21*** (-0.28)
Pronoun (PRP)	its	0.30*** (-0.12)			as	0.93*** (-0.16)
	it	-0.07 (-0.05)			of	1.21*** (-0.28)
	these	-0.1 (-0.06)			out	-0.01 (-0.07)
	this	-0.14 (-0.21)			in	-0.07 (-0.21)
	those	-0.14*** (-0.05)			for	-0.16 (-0.19)
	your	-0.31*** (-0.05)			with	-0.22 (-0.15)
	you	-0.66*** (-0.06)			into	-0.09* (-0.05)
	us	-0.77*** (-0.07)			per	-0.15*** (-0.05)
	we	-0.79*** (-0.06)			if	-0.65*** (-0.15)
	our	-0.92*** (-0.06)		Cardinal Digit (CD)	1	-0.19** (0.08)
	those	-0.14*** (-0.05)			2	-0.14** (0.07)
Preposition (IN)	at	0.04 (-0.15)			3	-0.58*** (0.10)
	on	0.04 (-0.09)			4	-0.48*** (0.10)
	against	0.10** (-0.05)			5	-0.52*** (-0.10)
	unless	0.10* (-0.06)			6	-0.51*** (0.10)
	except	0.12** (-0.06)			7	-0.38*** (0.07)
	upon	0.18* (-0.09)			8	-0.27*** (0.07)
	while	0.21* (-0.12)			9	-0.18*** (0.06)
	until	0.26* (-0.13)			10	-0.46*** (0.11)
	after	0.28** (-0.13)			30	0.11* (0.07)
	by	0.33* (-0.2)			60	-0.01 (0.10)

Note: $*p < 0.05$, $**p < 0.01$, $***p< 0.001$, ns: insignificant at the 0.05 level by POS tag. Coordinating Conjunction (CC), Pronoun (PRP), Preposition (IN), Cardinal Digit (CD) individual words are almost the same direction with their POS tag direction. In CC, as previous paper that "and" is a structural ambiguous word. In PRP, except "its", pronouns can specify the subject in sentence. So, pronouns reduce ambiguity then decrease the complaint ratio. Preposition (IN) cause syntactic ambiguity then increase the complaint ratio (Hindle and Rooth, 1993). However, "per" is usually used by "per day", "per person". It specifies the scope that reduces ambiguity. Moreover, "if" is interpreted in meaning analysis. Cardinal digit (CD) is specific numbers, so they tended to reduce ambiguity then reduce ambiguity.

4.5. Meanings Analysis

We interpret the individual words by their meanings. "Must" and "will" that used to specify the sentence and more exactly are negatively significant. "injury", "damage" is positively significant. Boardman (2005) said court's interpretation can be different from insurer's intent. And they gave an example word for "property damage". In addition, in an expression "as damage", if damage is used independently, it becomes ambiguous because it is unknown what kind of damage it is (Abraham, 1996). "Exclusion", "if", "include", "least" are negatively significant effect. Sentences that contain these words state the range to which the insurance company can compensate customers.

However, there has a limitation in this method, whether the words that can be used abstractly or specifically in the context of the sentence are different. Likewise, how ambiguous the meaning of other words can vary from sentence to sentence. In this study, there is a limitation because it did not specifically analyze how words were used in a sentence to sentence but analyzed them by word frequency. Nevertheless, the results of individual words classified into the POS category were relatively consistent. Even if interpreting by meanings, words that could increase ambiguity tended to have a high complaint ratio, and words that could lower ambiguity tended to have a low complaint ratio.

Ⅴ. Conclusion

Worldwide, the government's efforts to digitize are accelerating more than ever to realize public interest and increase public trust. The trend of digitalization affects individual industries as well as the direction of government. In particular, it helps the insurance industry prepare for the future by identifying strategic strengths and potential threats through data sharing and utilization.

In this sense, our research responds to the efforts of change within the insurance industry and the efforts of digital government to leverage data. In other words, in this study, by analyzing insurance terms through digitization, consumers will be able to avoid the consequences of choosing services they do not want by increasing the fitting with the services they want. Moreover, by companies' side, the opportunity is given to provide high-quality products in the industry by fitting the consumers' needs better and achieving consumer satisfaction.

Consequently, this paper verifies whether the language used in the contract affects customer dissatisfaction and whether the increase or decrease of complaint document is determined by the degree of linguistic factors. In particular, we identify whether firms' unobserved heterogeneous effects exist in contract documents. Further, to estimate the impact of ambiguity on consumers' dissatisfaction, we assess the contracts in three perspectives: overall ambiguity, word class, individual words. The results indicate firm-specific contracts affect the dissatisfaction of consumers. In terms of ambiguity, more clear contracts, frequent uses of the word classes that clarify personal pronouns, relatives, and cardinal digits are more likely to reduce the complaints. We interpret the difference in language usage affects the increase or decrease in complaints of insurance users. Therefore, specific words that clarify conditions or ranges reduce ambiguity; on the other hand, inaccurate words increase ambiguity, eventually affecting complaints.

The research findings suggest several contributions as follows. First, previous auto insurance complaint research analyzed general factors like legal cost or coverage (Carson, 2005; Doerpinghaus, 1991). However, in this study, we extend the related literature on text analysis concerning consumer satisfaction and firm performance. Second, we provide empirical evidence showing that ambiguity as a linguistic factor affects consumer dissatisfaction. In-stream of public data research, measuring insurance contract ambiguity is helpful for consumers as an evaluation tool for insurance contracts. Our findings and processing of public data sources of the insurance contract can improve the insurance industry by providing helpful business information. (Krishnamurthy et al., 2016) Third, this study helps to bridge the gap with empirical evidence for information asymmetry in both the supply and the demand sides in a contract context. Aside from most insurance studies that refer to demand-side adverse selection (Cohen et al., 2010; Einav et al., 2013), this study focuses on information asymmetry driven by the supply side. Since the service provider (insurance company) is an expert of legal words in contract compared to consumers, our paper contributes a new stream of research regarding supply-side adverse selection. As mention in the incomplete theory part, if the contract is incomplete and ambiguous, the possibility of reducing long-term customer-firm relationship increase. Depending on the incompleteness of the contract, customers' complaint to the company and information asymmetry increase. This perspective also assists the digital government's social purpose of sharing data and encouraging data utilization. Lastly, our study is a streamline of digital government's efforts to realize the social and economic goals of active data processing. Reducing ambiguity within contracts helps the completeness of contracts in the insurance policy and enhances information transparency by bringing monitoring effect. In addition, the economic value of data processing contributes to industry growth.

However, limitations resulting from placement of the sentences and words, the association of words and phrases within sentences, and the various heterogeneity that can arise in language context exist in our paper. Also, this paper cannot capture all of the ambiguity, in ambiguity literature, there are many kinds of ambiguity, but this paper captured only lexical ambiguity and structural ambiguity. Furthermore, the data is not clear-panel-data because in-state insurance government sites do not provide the data for all years equally. Some sites provide recent 3 years, or other sites provide 2005 ~ 2017 year data. Each company's insurance contract data also have limitations that governments do not provide the company's insurance contract by states and years. Nevertheless, insurance contracts do not have significant differences by state and year.

For future research, we plan to extend the data into precise panel data and make variables that can capture another ambiguity. Moreover, we extend to analyze not only auto insurance contracts but also other areas' insurance contracts.

Overall, our study expects to make theoretical contributions to academia and provide practical guidelines to the industry that can be used for consumer care and qualified service. Moreover, we give suggestions to policymakers that the findings can be used for better supervision of insurance policy.

<References>

[1] Abraham, K. S. (1996). A theory of insurance contract interpretation. *Michigan Law Review, 95*(3), 531-569.

[2] Andrew, N. M. (2019). *2018 complaint rankings: Accident & health: Auto insurance*. State of Connecticut: Insurance Department.

[3] Berry, D. M. (2003). *From contract drafting to software specification: Linguistic sources of ambiguity. A Handbook Version*. Citeseer.

[4] Bird, B. (2006). NLTK: The natural language toolkit. *Proceedings of the COLING/ACL Interactive Presentation Sessions*, Association for Computational Linguistics, 69-72.

[5] Boardman, M. E. (2005). Contra proferentem: The allure of ambiguous boilerplate. *Michgan Law Review., 104*, 1105-1126.

[6] Britton, B. K. (1978). Methods & designs: Lexical ambiguity of words used in English text. *Behavior Research Methods & Instrumentation, 10*(1), 1-7.

[7] Burton, S. (2018). The case for plain-language contracts want to do deals faster and increase customer satisfaction? Start by stripping out the legalese. *Harvard Business Review, 96*(1), 134-140.

[8] Carson, J. M., McCullough, K., and Russell, D. T. (2005). Complaint ratio and property-casualty insurer characteristics. *Journal of Insurance Issues, 28*(2), 151-166.

[9] Ceccato, M., Kiyavitskaya, N., Zeni, N., Mich, L., and Berry, D. M. (2004). *Ambiguity identification and measurement in natural language texts*. University of Trento, Technical Report DIT-04-111. (Unpublished)

[10] Coase, R. H. (1937). *The nature of the firm: Origins, evolution, and development*. Belknap Press.

[11] Cohen, A., and Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance, 77*(1), 39-84.

[12] Doerpinghuas, H. I. (1991). An analysis of complaint data in the automobile insurance industry. *The Journal of Risk and Insurance, 58*(1), 120-127.

[13] Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review, 103*(1), 178-219.

[14] Garrett. (1970). *Does ambiguity complicate the perception of sentences? Advances in psycholinguistics*. Amsterdam: North Holland Publishing.

[15] Grossman, S. F., and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy, 94*(4), 691-719.

[16] Hart, O. (1995). *Firms, contract, and financial structure*. Oxford University Press.

[17] Hart, O. (2017). Incomplete contract and control. *American Economic Review, 107*(7), 1731-52.

[18] Hart, O., and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy, 98*(6), 1119-58.

[19] Hindle, D., and Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics, 19*(1), 103-120.

[20] Karl, J. B., and Wells, B. (2016). Improving perceptions of the insurance industry: The influence of insurance professionals. *Risk Management and Insurance Review, 19*(1), 147-166.

[21] Kim, H., and Yang, S. J. (2019). A study on the consumer complaint of the insurance service-focused on the automobile insurance service. *Financial Planning Review, 12*(2), 1-33.

[22] Kim, J. (2012). The regulations on the mis-selling of insurance vehicles. *Chungnam Law Review*, 115-171.

[23] Knutsen, E. S. (2010). Auto insurance as social

contract: Solving automobile insurance coverage disputes through a public regulatory framework. *Alberta Law Review*, 715-750.

[24] Krishnamurthy, R., and Awazu, Y. (2016). Liberating data for public value: The case of Data. gov. *International Journal of Information Management, 36*(4), 668-672.

[25] Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313-330.

[26] Massey, A. K. (2014). Identifying and classifying ambiguity for regulatory requirements. *2014 IEEE 22nd International Requirements Engineering Conference(RE).*

[27] McKinsey. (2017, March). Digital disruption in insurance: Cutting through the noise. *Digital McKinsey*, https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/time%20for%20insurance%20companies%20to%20face%20digital%20reality/digital-disruption-in-insurance.ashx

[28] Ostrager, B. R., and Newman, T. R. (2018). *Handbook on insurance coverage disputes* (9th ed.). New York: Wolters Kluwer.

[29] Powell, L. (2017). Big data and regulation in the insurance industry. *Working Paper*.

[30] Singh, V. K., Mukherjee, M., and Mehta, G. K. (2011). Sentiment and mood analysis of weblogs using POS tagging based approach. *International Conference on Contemporary Computing, IC3 2011: Contemporary Computing*, 313-324.

[31] Spruit, M., and Ferati, D. (2020). Text mining business contract documents: Applied data science in Finance. *International Journal of Business Intelligence Research, 11*(2), 28-46.

[32] Torbert, P. (2014). A study of the risks of contract ambiguity. coase-sandor institute for law & economics. *Working Paper*, No. 686.

[33] Wang, T., Kannan, K. N., and Ulmer, J. R. (2013). The association between the disclosure and the realization of information security risk factors. *Information Systems Research, 24*(2), 201-497.

[34] Wells, B. P., and Stafford, M. R. (1995). Service quality in the insurance industry. *Journal of Insurance Regulation, 13*(4), 462-477.

[35] World Bank Blogs. (2021, March). *Digitalization and data can vastly improve public service delivery for citizens*. https://blogs.worldbank.org/europeandcentralasia/digitalization-and-data-can-vastly-improve-public-service-delivery-citizens

◆ About the Authors ◆

Jeongkwon Seo

Jeongkwon Seo is a Master student at the College of Business, Korea Advanced Institute of Science and Technology. His research interests include extracting meaningful information by text mining and applying in Information system and finance area.

Woojin Yang

Woojin Yang is a Ph.D. student at the College of Business, Korea Advanced Institute of Science and Technology. His research interests include the societal impact of platform business and the user behavior on blockchain-based platform.

Hyejin Mun

Hyejin Mun is a Ph.D. student at the College of Business, Korea Advanced Institute of Science and Technology. Her research interest includes the diverse impact of external shock on business and the impact of linguistic features.

Chul Ho Lee

Chul Ho Lee is an assistant professor of School of Business and Technology Management at KAIST. He obtained his Ph.D. in Management Science at the University of Texas at Dallas. He worked as a faculty member at Harbin Institute of Technology in China and Xavier University in US. He published many influential papers into decent journals such as Information Systems Research, Computers in Human Behavior, Journal of Knowledge Management, and so on. His research covers platforms, fintech, and decentralized finance.

Submitted: June 29, 2021; 1st Revision: September 2, 2021; Accepted: September 27, 2021