A DISCOURSE ANALYSIS THROUGH CORPUS LINGUISTICS ON THE CASE OF POPULAR SCIENCE BOOK *THE GRAND DESIGN*

Rosita Masfirotul Uyun
Universitas Airlangga
Surabaya
uyunarose92@gmail.com

Abstract

This study aims to discover the theme of popular physics book, The Grand Design, through corpus linguistics and discourse analysis as well as how the authors of the book describe the field of study which in fact is not their main focus to work on. In this case, the authors are physicists and cosmologists who attempt to use quantum to interpret the history if the universe. There were two terms, quantum and cosmology, which would be examined to gain the number of appearance in the book. A computer-based corpus software, AntConc v 3.4.3w and chi square test were employed to gain the hit numbers and the significant value of the possible different number appeared. This result would answer the question of what the theme of the book is. The procedure of corpus analysis, which raw text is directly analyzed, was used along with collocation and concordance analysis. After the data was gained from collocation and concordance analysis, discourse analysis was applied to obtain the profound findings of how the authors assert their opinion of quantum in the study of the cosmology. This paper resulted that quantum has more hit numbers than cosmology. In addition, it was found that the theory proposed in this book was quite inadequate in a certain circumstance.

Keywords: quantum, cosmology, corpus linguistics, discourse analysis.

Abstrak

Penelitian ini bertujuan untuk menemukan tema buku sains popular, The Grand Design, melalui korpus linguistik dan analisis wacana. Selain itu, penelitian ini berupaya untuk mengungkap cara penulis buku tersebut mendeskripsikan sebuah bidang kajian yang pada dasarnya bukan fokus kajian mereka. Dalam hal ini, penulis buku adalah fisikawan dan kosmologis yang berusaha untuk menggunakan teori kuantum untuk menjelaskan sejarah terbentuknya alam semesta. Terdapat dua istilah, quantum dan cosmology yang ada di buku dan akan diteliti secara linguistik. Sebuah perangkat lunak korpus bernama AntConc v 3.4.3w akan digunakan untuk mencari berapa kali dua istilah tersebut muncul dalam buku dan hasilnya akan dianalisis dengan menggunakan chi-square test untuk diketahui signifikansi perbedaan antara dua istilah tersebut. Data juga akan dianalisis dengan cara melihat kolokasi dan konkordansi, untuk selanjutnya diproses dengan menggunakan analisis wacana. Hasil menunjukkan bahwa quantum memiliki jauh lebih kemunculan dibandingkan dengan cosmology.

Di samping itu, ditemukan pula dalam buku tersebut bahwa terdapat teori yang belum cukup memadai untuk digunakan dalam situasi tertentu.

Kata kunci: quantum, cosmology, korpus linguistik, analisis wacana

1.   Background

The Grand Design was one of the most popular science books published in 2010. This book was authored by two incredibly bright physicists and cosmologists, Stephen Hawking and Leonard Mlodinow. Stephen Hawking is a British physicist and cosmologist who had been appointed the same position as Isaac Newton a few decades ago, as a Lucasian Professor of Mathematics in the University of Cambridge, England. When he was granted the professorship in 1979, he thought that he would only fill the gap in the position before the institution found the adequate candidate as he would not live longer (Mialet, 2003). Despite that, he was undoubtedly the most famous Lucasian Professor of Mathematics since Newton (Mialet, 2003: 426). In the autumn of 1962, he began to find difficulties in doing easy movement tasks, such as tying his shoes and having problem of talking (Ferguson, 2011). It kept getting worse in his third year at Oxford until he was diagnosed as having amyotrophic lateral sclerosis (ALS) when just turned twenty-one (Ferguson, 2011). In the premiere of 2015, a film about love story of Hawking and his wife was released. It was nominated in some number of categories in Academy Award 2015 and was awarded by the Best Performance of Actor in Leading Role in Academy Award 2015. He was not alone when he worked on *The Grand Design* as Leonard Mlodinow, an American physicist, author and screenwriter collaborated with him.

The book has eight chapters which each of them is connected to each other from a very brief history of the universe creation theories to how human is supposed to see the universe as being created by chance without the work of God. It has gained both positive and negative responses when it was just published. Hawking's colleague in his early career as a cosmologist, Roger Penrose, surprisingly had less good impression towards the book. In Financial Times (2010), he wrote that he doubted the way of the book was shaped for the public consumption was quite inadequate. He also called Hawking's philosophical standpoint as being strange-sounding. Another popular physicist who also wrote some best-selling science and religion books, Paul Davies, was also on the same boat as Roger Penrose. He in the Guardian (2010) wrote that the idea of which universe coming from a coincidental event and being accepted as a given was as unexplained as the idea of God.

In spite of the negative opinions of *The Grand Design*, the positive appreciation was raised from popular atheist biologist, Richard Dawkins. He told the Economist (2010) that he agreed to what Hawking proposed in his book, regarding to the idea of which the universe was created without God. Another positive opinion came from popular cosmologist and author Lawrence M. Krauss. In the Wall Street Journal (2010), he implied that what was written in the book was the data that proved the idea of God sooner or later disappears.

A few glimpse of the book becomes the interest of the researcher to conduct an investigation towards this book in the sight of linguistics. This study was conducted by applying frequency analysis, collocation and concordance analysis and in the end to draw well-defined perception, discourse analysis was used. By using a corpus software, *AntConc v 3.4.3w*, the theme of the book could be seen. The attempt was made to

sort the words which could be the field of science in particular, since the book has been labeled as popular science genre. There were two words emerging, which were *cosmology* and *quantum*. Both are technical terms that are rarely used in common text as Sinclair (2003) briefly explained that many of the infrequent words have constrained meaning, which are used only when people want to deliver an information to certain groups of people in certain situation. The reason of selecting both terms was rather subjective, because after the researcher read the book, the authors seemed to have tendency in both fields. Observing that the word *universe* also came into view repeatedly, it was important to consider whether it would also be included in the investigation. The idea, nevertheless, was dropped as it was in fact involved in cosmology and quantum discussion. In addition, it was the universe that became the main topic of this book. Hawking and Mlodinow (2010: 9-10) stated that:

> *"To understand the universe at the deepest level, we need to know not only how the universe behaves, but why. Why is there something rather than nothing? Why do we exist? Why this particular set of laws and not some other?"*

Unpredictably, the result exhibited in this paper that *quantum* has the more hit numbers of appearance in the book than *cosmology*. Though *quantum* appeared more than *cosmology*, chi square test was applied to know whether the result was significant. Later, in the methods, there will be more explanation of how data was gained. After the data was gained by using corpus, it was the interpretation that was important. Hence, the chi square test and Discourse Analysis was employed to figure out what could be inferred from the data in objective way, particularly how the authors who are actually physicists and cosmologists portrayed *quantum* in

their book. It is important to know how they describe a field which is decidedly different from the field they actually work on.

To be more specific, quantum focuses on the electromagnetic phenomena which the feature is several hundred times more microscopic than the proton (Peskin and Schroeder, 1995). On the other hand, Liddle (2003) described that the cosmological principle is both powerful and simple, as it becomes the keystone of which the beginning of the universe is being studied. Assuming the noteworthiness of the opinion presented by cosmologists towards the field in which they do not directly build up, in this case quantum, this paper attempted to reveal the frame of mind that is developed by the author of *The Grand Design* towards quantum.

To put in brief, this paper is to explore how a popular science book is analyzed in the view of language. Firstly, it is to reveal the theme of the book by using a corpus software since there is an unconventional point in the book. After the theme is displayed by a certain term, in this case *quantum*, it is used to show how the notion of *quantum* is depicted and placed within the text. To solve the matters of theme, this study employed frequency analysis and its significance, while to explicate the *quantum* portrayal, the concordance and discourse analysis were applied.

2. Literature Review

McEnery and Hardie (2012) explained that corpora represent the language produced in any mode, which is gathered and later used to indicate a fact of which the researcher is to seek for. Corpora provide the data that can be used as a standard reference to measure the language being investigated (Baker, 2006). It, for instance, allows the researcher to investigate a grammatical analysis in large-scale (*ibid.*). Then, it is necessary to match the research questions with the corpora (McEnery & Hardie, 2012).

The link between corpus linguistics and discourse analysis have become the concern of linguists. Flowerdew (2012) indicated that corpus linguistics and discourse analysis become the complex synergy of methods, approaches and tools, which the corpus linguistics now plays a central role in discourse analysis. Baker (2006) stated that the researcher's cognitive bias could be controlled, because the corpus limits the data we need. In discourse analysis, corpus linguistics is beneficial to help researcher to comprehend profoundly the subtle meaning inside the sentence which its phrase or grammatical construction could be investigated (Baker, 2006). Corpora presents the data of language produced in various ways in different time, such as web-based corpora called Corpus of Contemporary American English (COCA) and Corpus of Historical American English (COHA), British National Corpus (BNC) and many others which allow researcher to search for the data based on certain time. For this reason, discourse analysis can also gain the advantages in seeing the time changing in a discourse. This study nevertheless was not in attempt to see some pattern changing in a language, but practically investigated how a word in specific text was described.

In corpus linguistics, researcher is offered by the utilization that valuable to draw an analysis and interpretation through concordances and collocates. Since this paper was to examine lexical items and its notion within the context in the book, the analysis was based on the corpus-driven study, which the raw text was directly processed so that the pattern could be observed (Sinclair, 2004). This paper then employs concordance analysis to give detail information of the surrounding word of the central word, in this case *quantum* (Sinclair, 1991). Baker (2006) summarized that concordances assist researcher to see the semantic preference inside a specific pattern of the terms analyzed. McEnery and Hardie (2012) referred a technique called collocation-via-concordance which the word analyzed were independent of the statistical significance testing and were investigated by the items and patterns that occur repetitively.

A study conducted by Baker, Gabrielatos and McEnery (2013) presented that through concordances, the term *devout Muslim* was related to broad set of meanings, such as being told that a devout Muslim was criminally hypocritical, tended to have unacceptable behavior, was seen as outstanding when he/ she was described as normal and some other cases, a devout Muslim was viewed as a 'good' Muslim and Muslim who had suffered gained benefits through this term, as the sympathy for them was raised. The study also displayed the words that co-occurred frequently with *Islam* was *terror*. It was also proven by some other related terms such as *Islamic, Islamism, Islamist* and *Islamists*, which also happened to co-occur in long span with the terms *terror, terrorism, terrorist* and *terrorists* within the corpus text, especially after 9/11 (Baker, *et.al.*, 2013). Another study conducted by Bleich, Stonebraker, Nisar and Abdelhamid (2015) also investigated the British news headlines in various newspapers regarded to the Muslim portrayal in 2001-2012 through corpus. Later, the headlines were categorized into positive and negative prosodies (Bleich, *et.al.*, 2015). It was found that *muslims* are consistently depicted more negatively than Jews and frequently more negatively than Christians (*ibid.*). Gabrielatos and Baker (2008), who also examined the British media representation of refugees, asylum seekers, immigrants and migrants (RASIM), found that a number of categories of representation were mainly negative. Their study employed corpus-based analysis which they investigated the keywords focusing directly on RASIM reference and then the keywords were qualitatively examined via detailed line

by line concordance analysis (Gabrielatos & Baker, 2008: 15). Gabrielatos and Baker (2008, 21) argued that the frequency of semantic/ discourse prosodies is much higher than that of individual collocation patterns that give rise to them. It means that collocations could probably give the insight of how a word is described within a text, but it is not as much as semantic/ discourse prosodies do.

Jaworska and Krishnamurthy (2012) also presented collocation analysis of *feminism* represented in British and German press in 1990-2009. Based on press corpora of both British and German, the discourse towards *feminism* during that period was rather negative. The high frequency of collocates such as *dead* and *post-* portrayed that the feminist movement was old-fashioned. There was nevertheless a difference in term of cultural context between British and German. British press described feminism as being related to sexuality, whereas in German press, feminism was viewed as a movement in academic fields and arts rather than in social-political circumstances.

A study conducted by Hardy and Colombini (2011) also investigated the positive prosody of *risk* through frequency analysis, concordance contextual analysis, collocational analysis and a variation of on distinctive-collexem analysis, since some previous studies mostly have shown the negative prosody and medical result of the word *risk*. By using COCA, it was found that *risk* was frequently mentioned in the academic text comparing to fiction, newspaper, spoken and magazine (Hardy & Colombini, 2011). According to Hardy and Colombini (2011), to find the positive prosody of *risk* was exceedingly difficult, but they still found it such as "…**those risks are worth taking**."

Tang and Rundblad (2015) investigated how media report materials in drinking water that could potentially harm human health through *Wordsmith* software. By comparing US and British media, they found that although both media used different terminology of contaminants, they were likely represented the contaminants negatively (Tang & Rundblad, 2015). According to the concordance analysis, both media typically expressed the causal relationship between the threat and the object of risk, as well as the co-occurring words which represented the causal relationship such as *effects, impacts, impact, causing, affecting* and *cause* (Tang & Rundblad, 2015). Along with them, the words like *harmful, toxic* and *hazardous* also explain the relationship (*ibid.*).

The previous studies which were explained in the preceding paragraphs indicated that examining the concordance lines could be sufficient to provide the more detailed sense of a word within texts. Sinclair (2003) argued that grammar and lexical choice should be used to detail the meaning of a text, though grammarians think that grammar explains the general rules of language whereas lexis merely explains the meaning of individual words and phrases. There are areas in grammar where generalization is not allowed, yet when the "rules" can be "broken", the only way to describe them is by using the lexical choices (Sinclair, 2003: 63). Halliday (2005, 185) divided the function of grammar into three. Grammar of every natural has the active function that becomes the human experience. Grammar of ever natural language is reflective as it is the enactment of interpersonal relationship. Grammar can create discourse. A scientific theory contrasts daily life in that certain aspects or components of human experience are construed in semiotic subsystem in a different way, in the fashion of opening them up to be observed, investigated and explained (Halliday, 2005: 194).

In this paper, the two aspects of grammar, proposed by Gee (2014), was employed as the tool to analyze the discourse of the book. Gee (2014) divided grammar into two aspects which the grammar 1 is a study of a grammar as we have known such as a set of clause, phrase and so on and the grammar

2 is the rule of which grammatical units are used to create the pattern *whos-doing-whats-within-Discourse*. This grammar 2 is set into a list of inquiry tool that could beneficially decipher the meaning inside the text. This set is called grammar interludes (Gee, 2014). This paper however only applied a term called integrating information to reveal the facts showing inside the word *quantum*, after the data of which the words co-occurred with *quantum* and the context surrounding the text were gained through collocates and concordances analysis. Gee (2011) explained that to communicate information in certain perspective, speakers could integrate or package clause. The information that speakers try to deliver hence could be gained if the grammatical units of a phrase or a clause are put in detail and are explored the information from the each breakdown unit (Gee, 2011).

> *"When I was reading my book, I discovered that scientists think that hornworm growth exhibits significant variation."*

Gee (2011) exemplified the integrating information in sentence (1) by composing both list and figure of the detailed meaning inside the sentence. It could be seen as follows that a single sentence has significant value of information. The list of the clauses could be collapsed into (Gee, 2011: 59):

a. Main Clause: I discovered that scientists think that hornworm growth exhibits significant variation
b. Subordinate Clause: While I was reading my textbook
c. Embedded Clause: That scientists think that hornworm growth exhibits significant variation
d. Embedded Clause: The hornworms exhibit significant variation
e. Nominalization: Hornworm growth (hornworms grow)
f. Nominalization: significant variation (something varies significantly)

By integrating information in sentences of which taken from the words co-occurred with the central word, in this case *quantum*, this paper would try to reveal the portrayal quantum.

3.   Research questions

This paper would try to answer these research questions:

a. What the book is about seen through the corpus analysis?
b. How the word *quantum* is positioned in this book?

4.   Methodology

As the data was a kind of document, they would be included as qualitative data (Creswell, 2013). In this paper however data collection was also in form of quantitative, as an instrument called *AntConc v 3.4.3w* would be used to draw information. Later in the discussion, the interpretation would be provided firstly by quantitative data to answer the first research question. The quantitative analysis employed chi-square test by Microsoft Excel 2010. Qualitative interpretation, which discourse analysis was used to extract further analysis, was then delivered to support the quantitative findings (Creswell, 2004).

The format of the book *The Grand Design* was actually in .pdf, which later converted into .txt, in order to be able to be uploaded to *AntConc v 3.4.3w.* After the file was uploaded, the corpus was set to search two terms, *quantum* and *cosmology*. To answer the first research question, the researcher only needed to enter the term and the software would immediately have revealed how many numbers that each term was appeared in the text. For each term possibly had different hit numbers, it was needed to have a standard value which was used to determine whether the different numbers showed significant diversity. This study therefore employed chi square test to uncover whether the higher hit numbers of a term had the more possibility to become the main theme of the book.

The corpus was also used to gather the data that reveal how quantum was placed in the book about universe – because quantum is an atomic study, whereas the universe is the main object study of the cosmology. When it comes to the description, we firstly need to see the co-occurring words and sort them thematically. The concordance was afterwards viewed to the context following the words in wider range. Sinclair (1991) explained that there are some ways to help researchers to select information in concordance which either by frequency of by form. In this paper, to obtain the specification of word classes which were built up near the central word, the concordance was set in alphabetical order to help the analysis process. Besides, to answer the research question clearly, concordance allows us to see a specific character of any length such as a word, part of a word, or even part of a phrase (McEnery & Hardie, 2012). After the data needed was all collected and assorted into different subject matter, some sentences from each subject matter were analyzed using the grammar interludes (Gee, 2014). To simplify the result, the analysis was composed in the form of table with detailed description in it.

5. Result and Discusion
5.1. Quantum or Cosmology

By the mere sight, one would think that *The Grand Design* was a popular science book discussing the what, how, when and why of the beginning of the universe. If it is the universe that is being talked about, then it is cosmology that has the professional authority to the study of the universe. Cosmology studies the structure of the universe by reconstructing the history of the universe through the relevant physics (Raine & Thomas, 2001). However, when one look closely into the book, how the authors favorably used the theory of quantum as an example to describe the universe could be seen. To paraphrase Hawking and Mlodinow

(2010), if the smallest unit in the world such as photons could depart from the starting point where it was discharged to the indefinite endpoint in indefinite time, the enormous universe would also have indefinite number of universe, they approximate the number by 10^{500} universes that appeared in indefinite time, immediately after the big bang was banged and only one of them was the universe we know now.

The Grand Design book has tickled the curiosity nerve of the researcher to find out whether the book is actually about quantum or cosmology or even both. *AntConc v 3.4.3w* was used to reveal the hit numbers between the term *quantum* and *cosmology*. The result was out of the researcher prediction as the book predominantly discusses the universe, but it was showed in the picture below that *quantum* reached 141 hit numbers, while *cosmology* only hit 13 respectively.
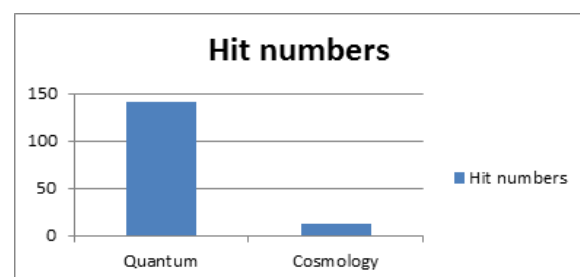

Figure 1. Hit numbers of *quantum* and *cosmology*

Nevertheless, we need to see whether the different numbers had significant value. The chi square test by Microsoft Excel 2010 showed that it was incredibly significant with p value = 0.00000000000000000000000006.

Conclusively, the book is about quantum and universe. Then the next question appeared is how quantum is placed within the discussion of enormous object in physics. It is needed to see the words that co-occurred with the term *quantum* in the book through collocates. It resulted that *quantum* frequently co-occurred with "physics", "theory", "field", "electrodynamics" and "chromodynamics".

Table 1. Frequency of the words co-occurring with *quantum*

Collocates	Frequency	Collocates	Frequency
physics	39	creation	3
theory	26	probability	2
theories	11	particle	2
field	7	model	2
event	4	jitters	2
electrodynamics	4	buckyball	2
chromodynamics	4	And	2
version	3	while	1
superposition	3	Vs	1
mechanics	3	versions	1
fluctuations	3	uncertainty	1

Those collocations later would be categorized into three semantic categories based on the similar description within texts. It is shown in the table as follows:

Table 2. Collocation category

Semantic Categories	Collocates
Change	- Probability
	- Superposition
	- Fluctuation
Difference	- Uncertainty
	- Version(s)
	- Vs (versus)
	- Physics
Particular field	- Chromodynamics
	- Electrodynamics
	- Theory(ies)
	- Mechanics

Change

"There will be a quantum **probability** *amplitude for every number of large space dimensions from zero to ten"*

Table 3. The detail of sentence (2)

	Information
There	- It is likely to show an abstract entity instead of a concrete one, such as a place. - It refers to a fact that is later introduced.
will be	It is the modal verb that shows the possibility of emerging fact in the future

	Information
a quantum probability amplitude	- There are three nouns which the head is "amplitude" with *quantum* and "probability" as the modifier and "a" as the indefinite article. - This phrase is the topic of the sentence. - Though the focus is supposed to be "amplitude", "probability" also tells us that the amplitude in quantum is not measured as indefinite numbers.
for every number of large space dimensions	- "every" shows particularity of something. - "large space dimensions: points out the dimensions that have large space. - This phrase represents the number of dimensions with large space.
from zero to ten	- It represents a measurement. - "from zero to ten" represents the number of dimensions with large space. - If it is to be related to the previous phrase, it can be inferred that there is no exact numbers of the measurement of the quantum amplitude for any number of dimensions.

"We are the product of quantum <u>fluctuations</u> in the very early universe"

Table 4. The detail of sentence (3)

	Information
We	- It is the main topic of the sentence. - It rather refers to the authors and the readers as it would be described later.
are	It shows a reference.
the product of the quantum fluctuation	- It explains that quantum fluctuations produce something. - It refers to a technical term in physics, which means that events in the early time when the universe was created.
in the very early universe	- It refers to a time when quantum fluctuation happened.

"In that view, the universe does not have just a single existence of history, but rather every possible version of the universe exists simultaneously in what is called a quantum superposition."

Table 5. The detail of sentence (4)

	Information
In that view	It refers to the subject of the previous sentence that is "the fundamental principle upon which our modern view of nature".
the universe	It refers to universe in literal meaning and is the main topic of the sentence
does not have just a single existence of history	It explains some numbers of histories of universe.
but rather every possible version of the universe	- It explains the opposite idea of the previous phrase - The universe has possibilities to have different version
exists simultaneously	It explains the previous phrase, which means that the different versions of the universe always exists
in what is called a quantum superposition	"superposition" refers to the technical term of

"Then came quantum uncertainty, curved space, quarks, strings and extra dimensions and the net result of their labor is 10^{500} universes, each with different laws, only one of which corresponds to the universe as we know it."

Table 6. The detail of sentence (5)

	Information
Then	It shows a sequence.
Came quantum, curved space, quarks, strings and extra dimensions and the net result of their labor is 10^{500} universes	- This is the main topic of the sentence, containing technical terms of physics - Those technical emerged.
each with different laws,	- "each" refers to the universes - Each universe has different laws
and only one of which corresponds to the universe as we know it.	- It shows the connection of the previous phrase and the next phrase. - One of the universe is what we know now.

Difference

"We therefore have to find quantum versions of all the laws of nature."

Table 7. The detail of sentence (6)

	Information
We	- It is the main topic of the sentence. - It seems to be the scientists, the book authors and the readers, yet when we see the next words, it could likely be the scientists.
therefore	- It is an adverb that describes a consequence. - As a consequence, the scientists should do something.
have to find	- Verb that expresses a compulsion. - The scientists have a compulsion to find something.
quantum versions of all the laws of nature	There are types of quantum governing the laws of nature which needs to be found.

Particular field

"Quantum physics is a new model of reality that gives us a picture of the universe."

Table 8. The detail of sentence

	Information
Quantum physics	- This is main topic of the sentence. - There is a field in physics called quantum.
is	It shows a reference.
a new model of reality	Reality has a new model
that gives us a picture of universe	This subordinate clause tells us that the new model of reality called quantum physics describes the universe.

"The quantum theory of electromagnetic field, called quantum electrodynamics, or QED for short, was developed in 1945 by Richard Feynman and others and has become a model for all quantum field theories."

Table 9. The detail of sentence

	Information
The quantum theory of electromagnetic field	- This is main topic of the sentence. - The quantum theory includes electromagnetics as its field.
called quantum electrodynamics	This is also the main topic which clarifies the previous phrase
or QED for short	This also the main topic, which shows that quantum electrodynamics can be abbreviated to QED
was developed in 1945	- The sentence is in the form of passives - This shows that somebody initiate an activity and makes changes.
by Richard Feynman and others	- "Richard Feynman" and "others" turned out to be the agent doing activity. - "Richard Feynman" becomes more important than the "others"
has become a model for all quantum field theories	- "model" does not refer to a prototype form or a form, but rather the standard. - The QED turns into an example for all quantum field theories.

"The strong force can be renormalized on its own in a theory called QCD, or chromodynamics."

Table 10. The detail of sentence

	Information
The strong force	This is the main topic. The force is strong
can be renormalized	This predicates explains that the main topic, "the strong force", can be in the unstable state which can also be normalized again
on its own in a theory called QCD	This phrase explains the predicates, meaning that the normalization can be done in a way which is called QCD
or chromodynamics.	This clarifies the abbreviation QCD.

"So though we don't yet have a complete quantum theory of gravity, we do know that the origin of the universe was a quantum event."

Table 11. The detail of sentence (10)

	Information
So	It shows a conclusion.
though	It points out that there is an opposite fact to what is said previously or afterwards. It also indicates that the following sentence was subordinate clause.
we	It refers to the readers, the scientists and the authors as it would be explained later.
don't yet have	It shows a lack of possession of something.
a complete quantum theory of gravity	A theory of gravity in quantum has been completed
We	It is the main topic of the main clause. It refers to the readers, the scientists of quantum and the authors. The authors try to involve the readers into the knowledge experienced by them and the quantum scientist. It also refers to the quantum scientists as the previous words discuss a theory in quantum field.
do know	It indicates an emphasis of knowing something

	Information
that the origin of the universe	This embedded clause shows how the universe began.
was	It shows reference in the past.
a quantum event	It shows an event in quantum field. A quantum event is the explanation of the origin of the universe

"In fact, Feynman once wrote, 'I think I can safely say that nobody understands quantum mechanics."

Table 12. The detail of sentence

	Information
In fact	It explains the truth of something.
Feynman	It is the main topic of the sentence. It refers to a quantum scientist that is explained in sentence (8).
once wrote	Some other day in the past, Feynman wrote something.
I	It refers to Feynman
think	It indicates a state action
I	It refers to Feynman
Can	It shows modality that presents a probable capability
safely say	In a careful way, the subject expresses something.
that nobody	The subject of the subordinate clause tells that there is no human who does something.
understands	It indicates a state action. There is no human who understands something.
quantum mechanics	It shows a field in quantum. Feynman wrote that there is no human who understands quantum mechanics.

6. Conclusion

The result indicates some numbers of interpretations. First of all, the authors in their books endeavor to explain how the universe began by using quantum theory, though theory is not accomplished yet. It could be seen in sentence (10) that in spite of the incomplete quantum theory of gravity, the authors insisted on their opinion which the universe began by quantum event.

"So though we don't yet have a complete quantum theory of gravity, we do know that the origin of the universe was a quantum event."

This is in line with sentence (6) and (7). Sentence (6) implies that there are types of quantum governing the laws of nature which needs to be found. Sentence (7) similarly suggests that a new model of reality called quantum physics accounts for the study of universe. It accordingly could be assumed that the authors evoke the untested hypothesis could be used to explain the universe.

"We therefore have to find quantum versions of all the laws of nature."
"Quantum physics is a new model of reality that gives us a picture of the universe."

Ellis (2013, 11) nevertheless argued that there are at least four criteria for a scientific theory, which a theory should have satisfactory structure, intrinsic explanatory power that the logical tightness and the scope of theory are required, extrinsic explanatory power that the connectedness to the rest of

science has to be obliged and observational and experimental support. Hence, the theory of which the universe is being governed by nature cannot completely be used in some circumstances as the last criterion, which a theory must have observational and experimental support, is not accomplished yet.

Quantum was also used to describe as having various fields, though quantum itself is one of major field in physics. Sentence (7), (8), (9) and (10) have the same suggestion that they explain how quantum is useful to describe the origin of universe. *Theory* is also included in *particular fields* as there is an area in quantum where it does not only discuss "theory" created by reasoning.

It could be then concluded from the result above that the book is in fact about the universe in the sight of quantum. It could be viewed by the frequency of term *quantum* emerging in the book with 141 times respectively. Besides, the authors also repeatedly used quantum theory when they defined the nature of the universe. They believed that quantum is the answer of the history of the universe.

This research could set off the other research topics topic related to corpus linguistics and discourse analysis, especially if the data is science texts rather than social or political texts. One of the research problem related to this study is the keyness. Culpeper (2009) explained that keyness describes "aboutness" of texts. He indicated that keywords can uncover the less easily noticeable features and lexical and grammatical patterns. The study resulted that some features associated with the character in the Shakespeare play could be revealed such as, the analysis of general adjectives and metaphorical color terms for Romeo and of plural common nouns for Mercutio (Culpeper, 2009).

References

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The Representative of Islam in the British Media Press.* Cambridge: Cambride University Press.

Bleich, E., Stonebraker, H., Nisar, H., & Abdelhamid, R. (2015). Media Portrayal of Minorities: Muslims in British Newspaper Headlines, 2011-2012. *Journal of Ethnic and Migration Studies, 41*(6), 942-962.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (4th ed.). California: SAGE Publications, Inc.

Culpeper, J. (2009). Keyness: Words, part-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics, 14*(1), 29-59.

Davies, P. (2010, September 4). Stephen Hawking's big bang gaps. *The Guardian*. Retrivied from http://www.theguardian.com/commentisfree/belief/2010/sep/04/stephen-hawking-big-bang-gap

Ellis, G. F. (2013). On the philosophy of cosmology. *Studies in History and Philosophy of Science part B: Studies in History and Philosophy of Modern Physics*(46), 5-23.

Ferguson, K. (2011). *Stephen Hawking: His life and work.* New York: Random House.

Flowerdew, L. (2012). Corpus-based discourse analysis. In J. P. Gee, & M. Hanford, *The Routledge Handbook of Discourse Analysis* (pp. 174-187). London: Routledge.

Gabrielatos, C., & Baker, P. (2008). Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Construction of Refugess and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics, 36*(1), 5-38.

Gee, J. P. (2011). *How to Do Discouse Analysis: A Toolkit.* London: Routledge.

Gee, J. P. (2014). *An Introduction to Discourse Analysis: Theory and Method* (4th ed.). London: Routledge.

Halliday, M. (2005). Things and relations: regrammaticising experience as technical knowledge. In J. R. Martin, & R. Veel, *Reading Science: Critical and functinal perspective on discourse of science* (pp. 185-236). London: Routledge.

Hardy, D. E., & Colombini, C. B. (2011). A genre, collocational and constructional analysis of RISK. *Internation Journal of Corpus Lingustics, 16*(4), 462-485.

Hawking, S. W., & Mlodinow, L. (2010). *The Grand Design.* London: Bantam Books.

Jawroska, S., & Krishnamurthy, R. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press, 1990-2009. *Discourse and Society, 23*(4), 401-431.

J. P. (2010, September 5). Another ungodly Squabble. *The Economist.* Retrieved from http://www.economist.com/blogs/babbage/2010/09/science_and_religion

Krauss, L. M. (2010, September 8). Our Spontaneous Universe. *The Wall Street Journal.* Retrieved from http://www.wsj.com/articles/SB10001424052748703946504575469653720549936

Liddle, A. (2003). *An Introduction to Modern Cosmology.* West Sussex: John Wiley & Sons.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Mialet, H. (2003). Is the end in sight for the Lucasian chair?: Stephen Hawking as Millenium Professor. In K. C. Knox, & R. Noakes, *From Newton to Hawking: A History of Cambridge University's Lucasian Professors of Mathematics.* Cambridge: Cambridge University Press.

Peskin, M. E., & Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory.* Colorado: Westview.

Penrose, R. (2010, September 4). The Grand Design. *The Financial Times.* Retrieved from http://www.ft.com/cms/s/2/bdf3ae28-b6e9-11df-b3dd-00144feabdc0.html#axzz1CSIgPlwa

Raine, D., & Thomas, T. (2001). *An Introduction to the Science of Cosmology.* Bristol: Institute of Physics Publishing.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, J. (2003). *Reading Concordances: an introduction.* London: Pearson Longman.

Sinclair, J. (2004). *Trust the Text.* London: Routledge.

Tang, C., & Rundblad, G. (2015). When Safe Means 'Dangerous': A Corpus Investigation of Risk Communication in the Media. *Applied Linguistics*, 1-23.