information

MDPI

Article

Topic Modeling for Analyzing Topic Manipulation Skills

Seok-Ju Hwang [1], Yoon-Kyoung Lee [2], Jong-Dae Kim [1,3], Chan-Young Park [1,3] and Yu-Seop Kim [1,3,*]

[1] Department of Convergence Software, Hallym University, Chuncheon-si 24252, Gangwon-do, Korea; tjrwn1121@gmail.com (S.-J.H.); kimjd@hallym.ac.kr (J.-D.K.); cypark@hallym.ac.kr (C.-Y.P.)
[2] Division of Speech Pathology and Audiology, Hallym University, Chuncheon-si 24252, Gangwon-do, Korea; ylee@hallym.ac.kr
[3] Bio-IT Center, Hallym University, Chuncheon-si 24252, Gangwon-do, Korea
* Correspondence: yskim01@hallym.ac.kr; Tel.: +82-10-2901-7043

Abstract: There are many ways to communicate with people, the most representative of which is a conversation. A smooth conversation should not only be written in a grammatically appropriate manner, but also deal with the subject of conversation; this is known as language ability. In the past, this ability has been evaluated by language analysis/therapy experts. However, this process is time-consuming and costly. In this study, the researchers developed a Hallym Systematic Analyzer of Korean language to automate the conversation analysis process traditionally conducted by language analysis/treatment experts. However, current morpheme analyzers or parsing analyzers can only evaluate certain elements of a conversation. Therefore, in this paper, we added the ability to analyze the topic manipulation skills (the number of topics and the rate of topic maintenance) using the existing Hallym Systematic Analyzer of Korean language. The purpose of this study was to utilize the topic modeling technique to automatically evaluate topic manipulation skills. By quantitatively evaluating the topic management capabilities that were previously evaluated in a conventional manner, it was possible to automatically analyze language ability in a wider range of aspects. The experimental results show that the automatic analysis methodology presented in this study achieved a very high level of correlation with language analysis/therapy professionals.

Keywords: LDA; Doc2Topic; topic modeling; topic manipulation skills analysis

check for updates

Citation: Hwang, S.-J.; Lee, Y.-K.; Kim, J.-D.; Park, C.-Y.; Kim, Y.-S. Topic Modeling for Analyzing Topic Manipulation Skills. *Information* **2021**, *12*, 359. https://doi.org/10.3390/info12090359

Academic Editor: Kostas Stefanidis

Received: 10 July 2021
Accepted: 24 August 2021
Published: 31 August 2021

1. Introduction

Conversation refers to interactive linguistic communication between two or more people [1]. Smooth communication plays an important role in maintaining smooth relationships with people, which can only be achieved with good conversational or linguistic skills [1,2]. Generally, a person's language ability grows explosively between the ages of four and six [3,4]. Subsequently, language ability grows gradually through a variety of experiences [5,6]. Language development is continuously conducted upon reaching the school age [7–11].

The degree of language development varies for each individual, and it is significantly delayed in cases of disability. This delay in language development causes many problems such as learning development disorders due to difficulties in communication. Therefore, it is necessary to identify the language development level at an early stage and to proceed with treatment [12–15]. Language therapists conduct analyses of language development levels; however, this process is time-consuming and requires a high level of expertise [16].

To address this problem, many studies have attempted automated analysis [17–21]. Among them, ref. [18] developed the Hallym Systematic Analyzer of Korean language that automatically analyzes Korean data using the methodology outlines in [17]. This system mainly analyzes the structural level of conversation using a morpheme analyzer and parsing analyzer. However, analysis of the subject matter related to the overall dialogue is not possible using this technique. Instead, this can be analyzed using an indicator of the subject operation ability. Topic manipulation skills refer to the ability to maintain a topic or present additional information during a conversation, as well as initiate a new topic [1,22,23]. If the topic operation ability is low, the context and situation of the entire conversation cannot be assessed, thus prohibiting normal conversation [24–26]. Therefore, language therapists also need to analyze various grammatical indicators for language development diagnosis. A previous study used Sent2Vec to estimate the degree of subject similarity across sentences in Korean dialogue [19].

In this study, we introduce a topic modeling technique (Latent Dirichlet Allocation: LDA, Doc2Topic) for the analysis of topic manipulation skills to automatically calculate the number of subjects and the rate of topic maintenance [27,28]. In addition, three Korean morphological analyzers (UTagger, Kkma and Okt) were used to prepare Korean conversation data, and their performance was compared.

According to the results of the experiment, the subject count measurement using LDA showed a correlation of 0.9765 with the results of the studies in [7,8] when analyzing morphology using UTagger. In addition, the rate of topic maintenance assessed by LDA achieved a correlation of 0.9353 with Okt. Therefore, LDA could better determine similar topics of relevance compared to Sent2Vec. Following this achievement, we aimed to enhance the system's applicability by adding an assessment of subject management capacity to the existing Korean automatic analysis program [18].

Section 2 of this study describes the relevant research. Section 3 presents the format of the data, along with some examples, including the procedure used for data creation. Section 4 evaluates various forms of analyzers and topic modeling techniques. Section 5 outlines the results of the experiment, and Section 6 provides the conclusions.

2. Related Works

The Hallym Systematic Analyzer of Korean language consists of the client and the server [18]. The client sends the conversation transcription file for analysis through the WEB. After its analysis using various indicators, the server provides the results of speaker analysis, the average of each indicator for the same age group as the speaker and the average contrast ratio of the speaker. The Hallym Systematic Analyzer of Korean language has a function that analyzes the grammatical ability of the speaker. Building on the existing system [17], we analyze the speaker's conversation ability more specifically by adding grammatical modalities, syntax structure, verbs/speech and the number of utterances per conversation turn. These existing systems provide a detailed analysis of conversation ability, but do not include indicators addressing topic manipulation skills [18]. A previous study ref. [19] presented a method using Sent2Vec to detect hot topics in sentences using minimal data. This methodology represented the main motivation for this study.

Ref. [8] looked at the development of conversation turn ability and topic manipulation skills at school age and after school age. This study was conducted for three categories: elementary school students, middle school students and high school students. To extract objective results, the study continued conversations on three topics ('family life', 'school life' and 'other/friends') that were familiar to the topics, through which they collected utterance data. Using the collected utterance data, the analysis was conducted by calculating the rate of topic initiation, the rate of topic maintenance and the rate of topic change of each group. According to the results of ref. [8], in terms of the number of topics, the rate of topic initiation and the rate of topic change, it was found that the high school group is lower than the middle school group, and the middle school group was lower than the elementary school group. In addition, in the rate of topic maintenance, the high school group was higher than the middle school group, and the middle school group was higher than the elementary school group.

Ref. [7] is a similar study to ref. [8], which conducted a study to confirm the topic manipulation skills of 1st, 3rd and 5th graders belonging to school age. In the case of the number of topics, the fifth-grade group was smaller than the third-grade group, and the third-grade group was smaller than the first-grade group. In the rate of topic maintenance, the fifth-grade group was higher than the third-grade group, and the third-grade group was higher than the first-grade group. The two studies show that language skills continue to develop from childhood to adolescence.

3. Data

Section 3 describes the procedure for creating utterance data and the structure of the data. In addition, by analyzing the data with various indicators, we check the degree of language development.

3.1. Data Collection

Data were collected from elementary, middle and high school students, and the process of obtaining ignition data was as follows [8,26]. Interviewers conduct conversations around three topics familiar to children ('family life', 'school life' and 'other/friends') to induce the natural speech of the interviewee. When the conversation begins with the theme of "family life," the interviewer continues to respond with a neutral response, such as "yes", "I see," to the interviewee's words to obtain a lot of conversation data, drawing the subject's words. If the subject stops in the middle of a conversation, the interviewee facilitates the conversation, allowing the interviewee to speak more. If the interviewee stops talking because there is nothing more to say, the interviewer changes to the topic of "school life" and proceeds as previously explained. After such an "other/friend" talk, it is transcribed and saved in a file. At this time, utterances that are not pragmatic, such as utterances that lose their meaning because the interviewee did not finish their words, utterances that do not fit the context, overlapped utterances with the intervention of an interviewer when the interviewee continues to speak, and self-talk, are excluded from transcription.

Existing studies using transcription files made in a similar format compare and analyze the degree of language development using various indicators such as dialogue sequence, number of utterances per dialogue sequence [7], grammatical morphemes [29] and oral language ability [30]. In this study, a total of 50 data were composed of 25 elementary school students, 15 middle school students and 10 high school students, and the number of topics and the rate of topic maintenance are analyzed using the topic modeling.

3.2. Data Examples

Table 1 shows a part of interviewee A's data. Turn means the sequence of conversations, and utterance means the number of utterances of interviewee A. Content to show the conversation with the interviewer and the interviewee. Interviewee A looks at the family-related pictures and thinks of the family and starts the story. In this case, it can be said that the child has started talking about the family well. Starting with the big topic of family, we can see that the second turn continue in a similar context of the first. At this time, it is confirmed that the interviewer responds to interviewee A's words and encourages him to continue speaking.

Table 1. Data Example.

	Turn	Utterance	Language	Contents
Interviewee	1	1	Korean	우선 그림 중에 생일이라는 그림이 있잖아요.
				Useon geurim junge saengiliraneun geurimi itjaayo
			English	First of all, there's a painting called birthday.
Interviewer			Korean	응 생일이라는 그림이 있었지.
				Eung saengiliraneun geurimi iteotji
			English	Yeah, there was a painting called birthday.

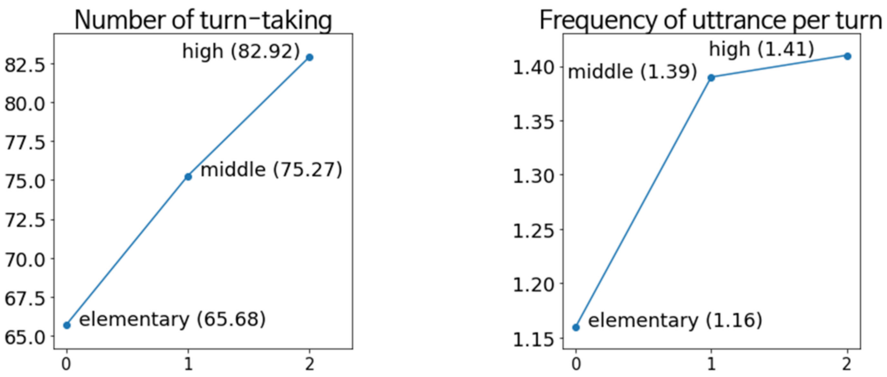
Table 1. *Cont.*

	Turn	Utterance	Language	Contents
Interviewee	2	2	Korean	최근에 동생 생일이었어요.
				Choegeune Dongsaeng Saengilieot Eoyo
			English	It was my brother's birthday recently.
Interviewer			Korean	아 동생 생일이었어.
				A Dongsaeng saengilieoteo
			English	Oh, it was your brother's birthday.
Interviewee	3	3	Korean	네.
				Ne
			English	yes
		4	Korean	근데 동생 놈이, 얘가 기브앤테이크가 안 돼요.
				Geunde dongsaeng nomi, Yaega gibeuaenteikeuga an dwaeyo
			English	But my younger brother, can't give and take.
interviewer			Korean	아 기브앤테이크가 안 되는구나.
				A gibeuaenteikeuga an doeneunguna
			English	Oh, give and take isn't working.

Since the interviewer simply responds to or promotes the interviewee's words to induce the child to speak more, we collected only the interviewee's utterances. We preprocess these utterance data and finally extract only the noun words. We then apply two topic modeling methods (LDA and Doc2Topic) to measure the number of topics and the rate of topic maintenance.

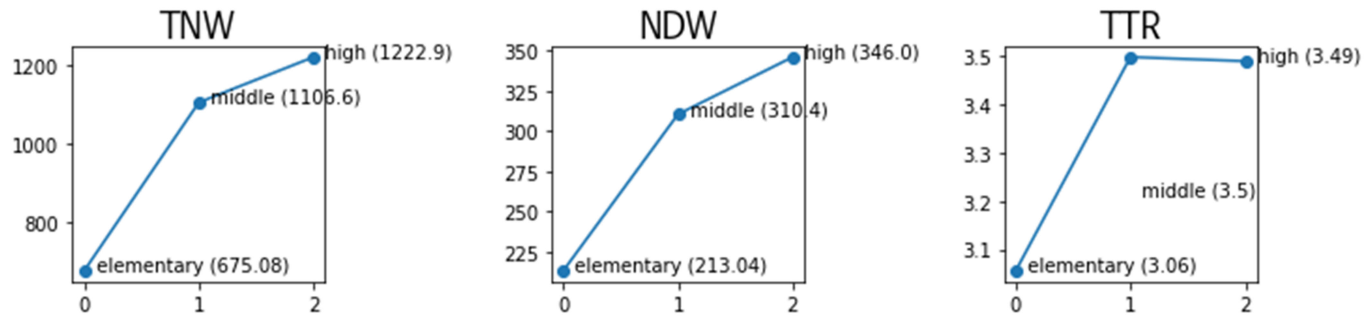
3.3. Data Analysis

In this study, the indicators measured in manual language diagnosis are automatically analyzed. In addition, the correlation between the automatic measurement results and the manual diagnosis results are compared with each other to analyze the correlation between the two. Figures 1 and 2 show the group average of each indicator. The x-axis represents the elementary school, middle school and high school students, and the y-axis represents the average of indicators.



(**a**) (**b**)

Figure 1. Indicators for conversational turn-taking skills. (**a**) Average number of turn-taking for elementary, middle and high school students; (**b**) Average frequency of utterance per turn for elementary, middle and high school students.



(**a**) (**b**) (**c**)

Figure 2. Indicators for word production skills. (**a**) Average TNW values for elementary, middle and high school students; (**b**) Average NDW values for elementary, middle and high school students; (**c**) Average TTR values for elementary, middle and high school students.

In Figure 1, (a) show the number of turn-taking of each class and (b) means frequency of utterance per turn. (a) refers to the total number of conversations exchanged with the interviewer and interviewee. (b) refers to the average number of utterances in one turn. These indicators can be calculated simply by counting the number of turns and utterances. We were able to obtain similar results from the study [7].

In Figure 2, (a) stands for Total Number of Words (TNW), (b) stands for Number of Different Words (NDW) and (c) stands for Type Token Ratio (TTR). TNW is the total number of morphemes uttered per interviewee, NDW is the size of morpheme vocabulary and TTR is TNW divided by NDW. This is the result of preprocessing the utterance data and measuring it through the result of morphological analysis. These indicators are important for measuring and diagnosing a speaker's lexical expression and speaking ability [8]. As the above results show similar results from the studies [31,32], it was confirmed that it is possible to evaluate the level of language development by age through the data used in this study.

4. Methods

In this work, we conduct experiments by dividing sentences into morphemes, which are the minimum semantic units. Sentences divided into morphemes are applied to Doc2Topic and LDA to analyze the number of topics and the rate of topic maintenance.

4.1. Morphological Analysis

A morpheme is the "minimum semantic unit" of language. At this point, meaning includes both vocabulary and grammatical meanings. Analyzing these morphemes is the most important and fundamental requirement in all natural language processing fields. Only after morphological analysis is completed, it can be applied to all natural language-related fields, including machine translation and natural language understanding systems, through parsing and semantic analysis. The morphological analyzer analyzes the given text in morpheme units and outputs it along with various lexical information including its part-of-speech.

In this study, experiments were conducted using Korean utterance data, so Kkma, Okt and UTagger were used similar to the results of expert analysis of experts for pretreating Korean. Kkma is a morphological analyzer written in Java and uses dynamic programming [33]. When analyzing text, it creates all possible candidates for the morphological analysis and arranges them in order that they are appropriate [19]. Okt is an open-source Korean language processor made on Twitter, which extracts index words with simple Korean processing.

UTagger carries out the identification of homograph words with a Korean morphological analyzer at the same time and uses Sejong TagSet for the part-of-speech tag. The homophone number system is based on Sejong and is generally consistent with the Standard Korean Language Dictionary (https://stdict.korean.go.kr/main/main.do, accessed on 1 August 2021) of the National Institute of the Korean Language (https://www.korean.go.kr/front_eng/main.do, accessed on 1 August 2021). It works by learning Sejong corpus and provides specialized functions of other domains with a technique called "user corpus". UTagger uses "user corpus" to learn new words, the usage of verbs and new contexts between two adjacent words in real time.

Table 2 shows examples of two sentences analyzed by an expert and analysis results of each type of analysis. Experts in the first sentence analyzed "positive" as a noun. UTagger and Kkma have analyzed the same analysis, but Okt analyzed "positive", not "positive". In the second sentence, experts analyzed "blog" as a noun. UTagge and Okt have analyzed the same as an expert, but Kkma analyzed "log" as a noun.

Table 2. Example of morphological analyzer.

Language	Sentence	Expert	UTagger	Kkma	Okt
Korean	긍정적인 요인도 있다. Geungjeongjeokin Yoindo Itda	긍정적/NNG 이/VCP ㄴ/ETD 요인/NNG 도/JX 있/VA 다/EFN ./SW	긍정적/NNG 이/VCP ㄴ/ETM 요인/NNG 도/JX 있/VA 다/EF ./SF	긍정적/NNG 이/VCP ㄴ/ETD 요인/NNG 도/JX 있/VV 다/EFN ./SF	긍정/Noun 적/Suffix 인/Josa 요인/Noun 도/Josa 있다/Adjective ./Punctuation
English	There is also a positive factor.				
Korean	오랜만에 블로그에 글을 올려 봅니다. Oraenma Beulrogeue Geuleul Olryeobopnida	오랜만/NNG 에/JKM 블로그/NNG 에/JKM 글/NNG 을/JKO 올리/VV 어/ECS 보/VV ㅂ니다/EFN ./SW	오랜만NNG 에JKB 블로그NNG 에JKB 글NNG 을JKO 올리VV 어EC 보VX ㅂ니다EF .SF	오랜만NNG 에JKM 블VV ㄹETD 로그NNG 에JKM 글NNG 을JKO 올리VV 어ECS 보VXV ㅂ니다EFN .SF	오랜만Noun 에Josa 블로그Noun 에Josa 글Noun 을Josa 올려Verb 봅Verb 니다Eomi .Punctuation
English	I posted a blog post after a long time				

Through this, it was found that there is a common point, but it may be different to the analysis. The data was pre-treated for each of the three analytical and experiments using Topic Modeling.

4.2. Topic Modeling

Topic modeling is one of the statistical models for discovering the abstract subject of a set of documents called topics in the field of machine learning and natural language processing. This is a text mining technique used to discover the hidden semantic structure of the text body. Among these topical modeling techniques, experiments are conducted on the number of topics and the rate of topic maintenance using LDA and Doc2Topic models.

4.2.1. LDA

LDA is a probability model for what topics exist on each document for a given document [15]. It is used to estimate both the distribution of words by topic and the distribution of topics by document, indicating how the topics in the document are distributed.

$$\mathrm{p}(\phi_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\phi_i|\beta) \prod_{d=1}^{D} p(\theta_d|\alpha) \left\{ \prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|\phi_{1:K}, z_{d,n}) \right\} \quad (1)$$

In the formula of (1), D means the total number of documents, K means the total number of topics and N means the number of words in the d-th document. $z_{d,z}$ assigns which topic the n-th word of the d-th document corresponds to. In addition, ϕ_i is the vector corresponding to the i-th topic, which represents the weight of the word in the i-th topic. Finally, θ_i is a vector representing the proportion of topics in the i-th document and is as long as the total number of topics K. Directly observable in the formula of LDA is the observable words from hyperparameter(α, β) and corpus, and the unobservable are z, φ and θ. z should have the maximum probability of combining the word distribution of the topic and the topic distribution of the document, to obtain this, it is necessary to find z, φ and θ, which make $p(z, \varphi, \theta \mid \mathrm{w})$ the maximum. To obtain this, the $p(\mathrm{w})$ value must be obtained, which refers to the probability of emergence of each w, taking into account the number of all cases of z, φ and θ. It is obtained by using the collapsed gibbs sampling technique.

Collapsed gibbs sampling is a method in which $p(z, \varphi, \theta \mid \mathrm{w})$ is automatically obtained by subtracting φ and θ from the calculation, using only z and w to obtain z. The expression of collapsed gibbs sampling is as shown in (2) and refers to the probability that the topic $z_{d,i}$ of the word d of the i-th document will be assigned to the j-th topic.

$$\mathrm{p}(z_{d,i} = \mathrm{j}|\mathrm{z}_{-i}, \mathrm{w}) = \frac{n_{d,k} + \alpha_i}{\sum_{i=1}^{K}(n_{d,i} + a_i)} \times \frac{v_{k,w_{d,n}} + \beta_{w_{d,n}}}{\sum_{j=1}^{V}(v_{k,j} + \beta_j)} = \mathrm{AB} \quad (2)$$

Expression (2)'s $z_{d,i}$ means a variable in the d-th document that indicates which topic the i-th word is assigned to, and z_{-i} refers to the topic information of all words except i-th. A represents the degree of association between the d-th document and the j-th topic, and B shows the degree of association between the n-th word and j-th topic of the d-th document. Molecular $n_{d,k}$ or $v_{k,w_{d,n}}$ comes out with a value of 0, which adds a hyperparameter (α, β) and a value to produce a smoothing-like effect. Repeat collapsed gibbs sampling 1000 to 10,000 times for all documents and all words, resulting in a convergence of the results, z, φ and θ values, and finally, all words are assigned a topic.

4.2.2. Doc2Topic

Doc2Topic is a neural topic modeling in LDA style [16]. Based on the skip-gram of the method of word2vec, instead of modeling the relationship between the central word and the context words, the relationship between the word id and the document id is modeled. Document id and word id are points in the same latent semantic space, and the dimension means the topic. This was successfully tested in a million documents with 200 themes, aiming to become an extensible alternative to LDA.

Figure 3 shows the structure of the Doc2Topic model. First, collect and preprocess the conversational data and apply a morpheme analyzer to extract only nouns. At this time, the nouns appear more than five times and create a document with only the number of occurrences greater than five times and encode the document whole number. Create document id using integer encoded values, and use this to create input_docs, input_tokens and label output. Then insert_docs and input_tokens as inputs to the model and enter the embedded layer. In the Embedding layer, change the shape to the number of topics dimension for values from input, and normalize the vector (L1), apply the RELU function, and change the shape again. Thus, input_docs and input_tokens are calculated(dot) and the values of the output are calculated as the binary correlation loss function on the last layer through the sigmoid function, and then optimized with the Adam algorithm. The optimization of this process over and over again completes the topic distribution of each sentence and word.

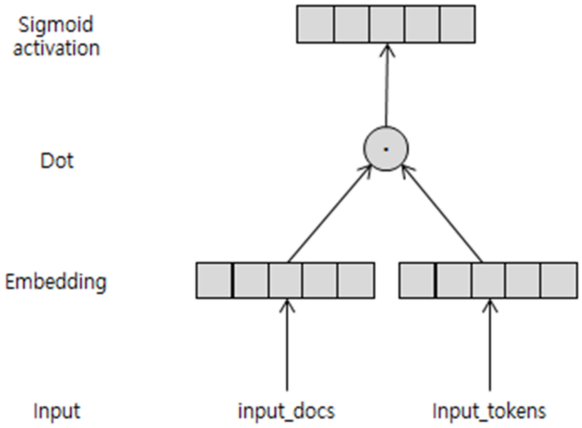


Figure 3. Doc2Topic Model Architecture.

The method of measuring the number of topics and the rate of topic maintenance is the same for both LDA and Doc2Top models. Set the parameter values from 3 to 30 to obtain and set the number of topics for each model. The number of topics with the highest coherence value was set to the model and measured by the number of topics, and the corresponding model was labeled to identify which topic each sentence belonged to. Through the labeling of each sentence in one data, if the label value of one sentence is 3 and the label value of the next sentence is equal to 3, it is determined that the topic is maintained and the rate of topic maintenance is measured accordingly.

5. Experiment

Both models have the same number of topics and methods for analyzing the rate of topic maintenance. In this study, it was assumed that the number of topics varied from 3 to 30, and the coherence value of the model was obtained for each number of topics. Among them, the number of topics is determined using the most consistent model, and the rate of topic maintenance is measured through each labeled sentence. In this work, some of the data used in ref. [8] are used, and the correlation between our model and the studies in ref. [7] and ref. [8] is analyzed.

Experimental Results

From the utterance data, only the utterance of the interviewee is collected except the words of the interviewer, and only the nouns are extracted using the morpheme analyzer Utager, Kkma and Okt. We then measure the number of topics and the rate of topic maintenance using the Doc2Topic model and the LDA model. Finally, the number of topics measured in ref. [7] and the resulting portion of the rate of topic maintenance and the correlation coefficients with the results tested for this study were identified, and the table summarizing them is as follows.

In Table 3, ref. [8] means a study targeting three groups of elementary, junior high and high school students, and ref. [7] means that the experiment was conducted in three groups of elementary school students in first, third and fifth grades. In Table 3, two things can be confirmed as a result of having a high correlation coefficient in both studies. First, the number of topics can be automated. The data was pre-processed with UTagger and applied to LDA, resulting in a number of topics. When the average of each group was calculated and the correlation coefficient with the study result was calculated, it was high [7,8]. Second, the rate of topic maintenance can be automated. This is the result of conducting a morphological analysis with Okt, applying it to the LDA, and measuring the rate of topic maintenance through labeling for each sentence, resulting in a high correlation coefficient through the results and the results from the research. [7,8]. Doc2Topic also indicated a high correlation overall when trying to measure the number of topics. After pre-preprocessed with Kkma, the average number of subjects applied to the Doc2Topic model and the correlation with the figure of ref. [8] was high. In the rate of topic maintenance section, the pre-proposed was conducted with Kkma, and when the subject retention was measured, it showed a high correlation with the results of the study [7]. As a result, both LDA and Doc2Topic models showed good correlation numbers, but for both studies, LDA showed better correlation numbers than Doc2Topic models.

Table 3. The Results of this Study.

	Morpheme Analyzer	**UTagger**		**Kkma**		**Okt**	
	Topic Modeling	**Doc2Topic**	**LDA**	**Doc2Topic**	**LDA**	**Doc2Topic**	**LDA**
Number of Topic	[8]	0.6579	0.9744	0.9609	−0.9860	0.8754	−0.9712
	[7]	0.9116	0.9789	−0.8796	−0.6883	0.6772	0.765
Rate of Topic Maintenance	[8]	−0.405	0.6578	−0.7909	−0.1677	−0.1527	0.9905
	[7]	−0.1534	−0.1945	−0.2867	0.944	−0.2866	0.8805

The number of utterance data used in this paper was conducted on the number of primary students, 15 middle school students and 10 high school students, using 10 data, and tested on the subjects and the rate of topic maintenance. The number of ignitions of each student varies from about 50 ignitions to the number of 120 ignitions. Even though it may not be the number of large amounts of data, it shows that the amount of data can be justified by showing the high correlation coefficient for both research.

6. Conclusions

In this study, we propose a method of analyzing the topic manipulation skills to analyze the language ability at a more different angle. Among them, methods to measure the number of topics and the rate of topic maintenance are implemented through LDA and Doc2Topic. We also wanted to confirm the difference in performance using a variety of Korean morphological analyzers. Among the topic modeling methods, LDA has shown a better performance, in some cases, the variation of that performance is severe. Among the Korean morphological analyzers, UTagger has shown relatively stable performance, but this was not to show consistent results. Therefore, the method of analyzing the topic manipulation skills based on topic modeling has the meaning that the evaluation in the qualitative realm has been transferred to a quantitative model, but it does not yet show predictable performance.

The biggest reason to see this aspect is because the definition of “maintenance of the topic” of this study is different from that of existing experts. Experts were recognized as the maintenance of the theme if the existing ‘theme’ was not seriously reflected even if a non-associated ignition was generated with the ‘theme’. However, when the system was found to be a bit shifted by the ‘theme’ of the top sentences, it was determined by the topic change.

Therefore, in the future, experts should be introduced into the system to determine the maintenance of the topic. Based on the topic modeling technique, we need to find a way to determine that the subject is being maintained more flexibly. This requires a way to determine the degree of association with the conversation topic. That is, it is necessary to be able to determine whether the current ignition is associated with the current topic, perhaps or is related to other topics. Future studies will raise the automation level of the speaker’s subject operational ability assessment to a higher level. Through this, it will enable the language ability of more children to enable the treatment of language development early.

Author Contributions: Conceptualization, Y.-S.K. and Y.-K.L.; data curation, Y.-K.L.; formal analysis, S.-J.H., J.-D.K. and C.-Y.P.; methodology, S.-J.H.; project administration, Y.-S.K.; resources, S.-J.H.; supervision, Y.-S.K.; validation, J.-D.K. and C.-Y.P.; writing—original draft, S.-J.H.; writing—review and editing, Y.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF) (NRF-2019S1A5A2A03052093).

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Heo, H.S.; Lee, Y.K. Conversational turn-taking and topic manipulation skills in conversations of school-age low-achievers in language learning. *Commun. Sci. Disord.* **2012**, *17*, 66–78.
2. Lee, H.J.; Kim, Y.T. Turn-Taking Characteristics of children with Specific Impairment and Normal Children. *Commun. Sci. Disord.* **2001**, *6*, 293–312.
3. Kim, Y.T. A Study on the accuracy of consonants in pre-school children using the picture consonant test. *Commun. Sci. Disord.* **1996**, *1*, 7–34.
4. Hong, J.H.; Pae, S.Y. The coda error patterns of children aged from 2 to 5. *Korean J. Commun. Disord.* **2002**, *7*, 294–304.
5. Cunningham, A.-E.; Keith, E.-S. What reading does for the mind. *Am. Educ.* **1998**, *22*, 8–17.
6. Shin, J.W.; Kim, J.Y. A Study of Korean Elementary School Students' Usage of English Vocabulary Learning Strategies. *J. Foreign Stud.* **2017**, *39*, 65–86.
7. Park, Y.J.; Choi, J.E.; Lee, Y.K. Development of topic management skills in conversation of school-aged children. *Commun. Sci. Disord.* **2017**, *22*, 25–34. [CrossRef]
8. Yang, Y.W.; Lee, Y.K.; Choi, J.E.; Yoon, J.-H. Development of conversational skills from late children to adolescent. *Commun. Sci. Disord.* **2018**, *23*, 270–278. [CrossRef]
9. Brinton, B.; Fujiki, M. Development of topic manipulation skills in discourse. *J. Speech Lang. Hear. Res.* **1984**, *27*, 350–358. [CrossRef] [PubMed]
10. Schober-Peterson, D.; Johnson, C.J. The performance of eight-to ten-year-olds on measures of conversational skilfulness. *First Lang.* **1993**, *13*, 249–269. [CrossRef]
11. Nippold, M.A. *Later Language Development: School-Age Children, Adolescents, and Young Adults*; PRO-ED, Inc.: Austin, TX, USA, 2016; p. 344.
12. Beilinson, J.S.; Olswang, L.B. Facilitating peer-group entry in kindergartners with impairments in social communication. *Lang. Speech Hear. Serv. Sch.* **2003**, 154–166. [CrossRef]
13. Lee, H.Y. Effects of Social Story Based Group Language Treatment on Basic Emotional Vocabulary Comprehension of School-Aged Children with Autism Spectrum Disorders. Master's Thesis, Ewha Womans University, Seoul, Korea, 2015.
14. Lee, J.H.; Lee, H.R. The relationship between evidentiality development and theory of mind in school-aged children. *Commun. Sci. Disord.* **2016**, *21*, 206–216. [CrossRef]
15. Richardson, K.; Klecan-Aker, J.S. Teaching pragmatics to language-learning disabled children: A treatment outcome study. *Child Lang. Teach. Ther.* **2000**, *16*, 23–42. [CrossRef]
16. Park, Y.R.; Choi, S.Y. The Effects of Conversational Skills Intervention Through Group Program on Conversational Turn-Taking and Topic Manipulation Skills of School-Age Children with Language Delay. *J. Speech Lang. Hear. Res.* **2019**, *28*, 115–128.
17. Choi, J.E.; Oh, B.D.; Heo, T.-S.; Kim, Y.-S. Automatic Analysis Service for Korean Speaking by Age. In Proceedings of the 30th Annual Conference on Human and Cognitive Language Technology, Seoul, Korea, 12–13 October 2018.
18. Hwang, S.J.; Oh, B.D.; Lee, Y.K.; Kim, Y.S. Hallym Systematic Analyzer of Korean (H-SAK) ver.1.0. In Proceedings of the Korean Institute of Information Scientists and Engineers, Pyeongchang, Korean, 18–20 October 2019.
19. Heo, T.S.; Lee, Y.K.; Kim, Y.S. Detection of Topic Changes in Child Speech Using Sent2Vec. In Proceedings of the 31st Annual Conference on Human & Cognitive Language Technology, Daejeon, Korean, 11–12 October 2019.
20. Finestack, L.H.; Rohwer, B.; Hilliard, L.; Abbeduto, L. Using Computerized Language Analysis to Evaluate Grammatical Skills. *Lang. Speech Hear. Serv. Sch.* **2020**, *51*, 184–204. [CrossRef]
21. Borden, G.A.; Watts, J.J. A computerized language analysis system. *Comput. Humanit.* **1971**, *5*, 129–141. [CrossRef]
22. Choi, J.E.; Lee, Y.K. Conversational turn-taking and topic manipulation skills of children with high-functioning autism spectrum disorders. *Commun. Sci. Disord.* **2013**, *18*, 12–23. [CrossRef]
23. Foster, S. The development of discourse topic skills by infants and young children. *Top. Lang. Disord.* **1985**, *5*, 31–45. [CrossRef]
24. Baines, E.; Howe, C. Discourse topic management and discussion skills in middle childhood: The effects of age and task. *First Lang.* **2010**, *30*, 508–534. [CrossRef]
25. Brinton, B.; Fujiki, M.; Powell, J.M. The ability of children with language impairment to manipulate topic in a structured task. *Lang. Speech Hear. Serv. Sch.* **1997**, *28*, 3–11. [CrossRef]
26. Schober-Peterson, D.; Johnson, C.J. Conversational topics of 4-year-olds. *J. Speech Lang. Hear. Res.* **1989**, *32*, 857–870. [CrossRef] [PubMed]
27. Oh, S.E.; Heo, T.S.; Lee, Y.K.; Kim, Y.S. Measurement of the number of topics in children's speech using LDA and Affinity propagation algorithm. *Korean Inst. Inf. Sci. Eng.* **2019**, *46*, 1403–1405.
28. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn.* **2003**, *3*, 993–1022.

29. Jung, K.H.; Pae, S.Y. The Use of Grammatical Morphemes of School-aged Children with Specific Language Impairment According to Discourse Type. *J. Speech Lang. Hear. Res.* **2010**, *19*, 161–176.
30. Heo, H.S.; Kwag, K.M.; Lee, Y.-K. The Relationship among the Reading and Writing Abilities and Oral Language Skills of School-Aged Low-Achievers in Language Learning. *Commun. Sci. Disord* **2011**, *16*, 23–33.
31. Lee, P.Y.; Kim, J.S. A Study on the Expressive Lexical Abilities of Elementary School Students. *J. CheongRam Korean Lang. Educ.* **2008**, *38*, 219–237.
32. Kim, Y.T.; Park, H.J.; Min, H.-K. School-Aged Children and Adults's Core Vocabularyfor the Development of an Augmentative andAlternative Communication Tool. *Korean J. Commun. Disord.* **2003**, *8*, 93–110.
33. Lee, D.J.; Yeon, J.H.; Hwang, I.B.; Lee, S.G. KKMA: A tool for utilizing Sejong corpus based on relational database. *J. KIISE Comput. Pract. Lett.* **2010**, *16*, 1046–1050.