Automatic Classification of the "De" Word Usage for Chinese as a Foreign Language

Jui-Feng Yeh* and Chan-Kun Yeh*

ABSTRACT

This paper proposed a word usage classification for "De" in Chinese as a secondary language by rule induction algorithm. Learning of Chinese characters and tone adaption are both essential and hard tasks for non-native speakers. The frequent terms, defined in morphosyntatic particle "De" with three characters {的, 得, 地}, is hard to learn for foreign learners due to the similar pronunciation and meaning. This investment illustrates a data-driven algorithm to classify the usages about the morphosyntatic particle "De" in Chinese learning. Rule induction is one of the most important techniques to learn the knowledge from data. Since regularities hidden in data are frequently expressed in terms of rules, rule induction is one of the fundamental tools for natural language processing and obtains a significant improvement in character selection. By the automatic rule induction process, 32 rules are adopted here to classify the character usage in morphosyntatic particle "De." According to the experimental results, we find the proposed method can provide good enough performance to classify the character usages for morphosyntatic particle "De."

Keywords: Rule Induction, Natural Language Processing, Secondary Language Learning, Classifier, Word Usage**.**

1. Introduction

To learn Chinese as a foreign or second language is to study of the Chinese languages by non-native speakers and new learners. Increasing interested peoples in China learning from those outside has led to a corresponding interest in the study of Chinese as their second language, the official languages of mainland China and Taiwan. However, the learning of Chinese both within and outside China is not a recent phenomenon. Westerners started learning different Chinese languages in the 16th century. Within China, Mandarin became the

---

* Department of Computer Science and Information Engineering, National Chiayi University, Chiayi city, Taiwan, ROC
  E-mail: Ralph@mail.ncyu.edu.tw

official language in early 20th century. According to the analysis of Summer Institute for Linguistics (SIL), there are near to seven thousands languages over the world nowadays. Among these languages, the top five languages are Chinese, English, Spanish, Bengali and Indian by their population sizes. As the first and second languages, Chinese occupies 14.8 percents populations to be the most used language. China's growing global influence has prompted a surge of interest in learning Mandarin Chinese as a foreign language (CFL), and this trend is expected to continue. Therefore, the population to learn Chinese as the second language is increasing in the latest decades (Simpson, 2000). Compared to the alphabetic language, Chinese is more complex and hard to understand for non-native speakers due to its several thousand characters and complicated sentence structures. Due the historical evolution of Chinese is deep and far, there are some word usage is susceptible to the corresponding allusions. Therefore, it is hard for the second language learners without the Chinese cultural background to understand, handle and use with skill the Chinese words very well. Actually, there are many whereas many computer-assisted learning tools have been developed for learning English, support for CFL learners is relatively sparse, especially in terms of tools designed to automatically evaluate learners' responses.

Computer technologies are used to assist in language learning, the so-called Computer-Assisted Language Learning (CALL), has been invested in the latest decades. An investigation was proposed to the adoption of information and communication technology (ICT) for teachers of Chinese as a foreign language (CFL) in US universities (Lin *et al.*, 2014). Yang (2011) emphasized an online situated language learning environment, for supporting the students, the teachers, and the teaching assistants (TAs) to communicate synchronously and asynchronously in and after class. Chen and Liu (2008) proposed a web-based synchronized multimedia lecture system based on WSML for the learners to learn Chinese as second language. They also compared the Web-CALL, IWiLL, and BRIX based systems for evaluating the proposed systems in Chinese learning/teaching (Chen & Liu, 2008). A user-centered design approach for learn Chinese as second language was invested for evaluating the web usability in (Huang *et al.*, 2010). Lu *et al.* (2014) suggested the curriculum content design in learning Chinese as a second language.

However, Chinese is rated as one of the most difficult languages to learn for people whose native language is English, together with Arabic, Japanese and Korean. There are many difficulties for foreigners to lean Chinese as their second language mainly caused by the special character set and tones in Chinese. Pronunciation cannot be obtained from its character directly. Although there are three aspects: text shape, pronunciation, meaning within one Chinese character. However, there are differences in pronunciation among the similar characters. Therefore, it is hard for the foreign learners to spell the correct Chinese words. For preventing the word segmentation error confusing the word boundaries, Bai *et al.* (2013) used

the inter-word spacing effects on the acquisition of new vocabulary for readers. One character with different pronunciations and meanings is hard to understand for non-native learners. Compared to other languages, the information of the Chinese character is overloaded. Besides, the number of the Chinese character is too large for a novice especially for the character with server usages. Tone is not easy to control in Chinese characters. The four tones are hard to enunciate for the non-native learners with a toneless source language. For example, the pitch trajectories for the secondary and third tones are one of the main obstacles for the learners. Accented pronunciation confuses the learners to obtain the standard. The pronunciation in Chinese is usually influenced by the speakers' own dialect, since the speaker has learned the dialects before they use the Mandarin. The usages of mandarin usually are affected by the dialects significantly such as Wu, Hokenese, Haka, and Cantonese. The complex structure of the Chinese character makes the hinder for nonnative learners. Reading and writing are main learning activities and they are cross validation for assessment of the achievements to use the Chinese characters. However, the complex structure and too many strokes make it more difficult to understand the reading and writing for learners. The flexible grammar rules in Chinese are not easy to learn for nonnative speakers. Confucius has described the Chinese as "a language without solid grammar (文無定法)"since two thousand years ago. The flexibility in syntax makes Chinese to be one of the most various languages. The rich rhetoric in Chinese make it is interesting and hard to understand the grammatical rules. The influence by ancient writings, the word usage is more complex in Chinese. That is to say, the literary language used in ancient China and preserved today for formal occasions or conspicuous display. Without the culture background, the foreigner learners are not able to obtain the meaning and pronunciation about word preciously.

The part of a word to which inflectional endings are attached, they are usually seen in alphabetical languages. Stem provides a good extension for word usage for language learners. However, the stems are hard to be obtained for non-native speaker, since the Chinese word with complex structure. The lexicon is hard to use for new learners. Actually, the design of Chinese lexicon aims at the user who is experienced in Chinese usage especially for the populations in home country. It is not friend for new learners. This makes it hard to study Chinese by oneself. For removing the barrier of learning Chinese as second language, more efforts are invested in Chinese character learning. Learning Chinese, which consists of more than ten thousands of characters composed of hundreds of basic writing units, presents such a challenge of orthographic learning for non-native speakers at the beginning stages of learning. A classroom was designed to extend previous research on how to support orthographic learning in (Chang *et al*., 2014). Chuang and Ku (2011) invested the effect of computer-based multimedia instruction with Chinese character recognition for foreign learners. Chen *et al*., (2013) proposed an approach for investigating the a radical-derived Chinese character

teaching strategy on enhancing Chinese as a Foreign Language (CFL) learners' Chinese orthographic awareness based on statistical data from the Chinese Orthography Database Explorer established and used as an auxiliary teaching tool. Hsiao *et al*. (2013) designed and developed a Chinese character handwriting diagnosis and remedial instruction (CHDRI) system to improve the CFL learners' ability in Chinese character writing. The CFL learners were given two tests based on the CHDRI system. One test focused on Chinese character handwriting to diagnose the CFL learners' errors in the stroke order and their knowledge of Chinese characters, while the other test focused on the spatial structure of Chinese characters (Hsiao *et al*., 2013). Looi *et al*. (2009) Explored interactional moves in a CSCL environment for Chinese language learning. Chang *et al*. (2012) presented approach for error diagnosis of Chinese sentences for Chinese as second language (CSL) learners. A penalized probabilistic First-Order Inductive Learning (pFOIL) algorithm is presented for error diagnosis of Chinese sentences. The pFOIL algorithm composed with three parts: inductive logic programming (ILP), First-Order Inductive Learning (FOIL), and a penalized log-likelihood function for error diagnosis (Chang *et al*., 2012). Chinese is a tonal language; tone and pronunciation acquisition also plays an essential role for CSL learners. There are some research efforts were made for listening and speaking diagnosis (Hao, 2012; Chu *et al*., 2014; Chun *et al*., 2015; Hsiao *et al*., 2015).

Since the learning for Chinese is not easy for non-native speakers. This drives us to the question what is the one to help the foreign learners. Indeed, the characters those are frequently used and mistake for each other usually confuse the foreign Chinese learners. The second language learners for Chinese usually are in the state of confusion about the usage of "De" (Jiang *et al*., 2012). Shi and Li (2002) analyzed the causal relationship between the establishment of classifier system and the grammatical issues of the particle "De". Yip and Rimmington (2004) described that "De" is required to be present in the relative clause as modifier contexts for Chinese as second language (CSL) learners. Waltraud (2012) analyzed the insubordinate subordinator "De" in Mandarin Chinese. Paul (2012) compared the difference of "De" in Chinese and French. Li (2012) also compared the usage between "De" in Chinese and "E" in Taiwanese. This paper invested an automatic rule induction algorithm for classification of the usages of the morphosyntactic particle"De." The confusing set about the morphosyntatic particle "De" is defined as the character set {的, 得, 地} in Chinese. Herein, the automatically classification about the morphosyntactic particle "De" is further defined as the process to decide which character is correct for using in Chinese. That is to say, we want to help the non-native learner to know which one is correct in the morphosyntactic particle "De" in Chinese.

This paper is organized as follows. Section 2 describes the rule induction algorithm used for classify the usage of morphosyntactic particle "De" in detal. In Section 3, we analyze the

performance in experimental results of the proposed methods. Finally, Section 4 will illustrate the findings and draw the conclusion of this paper.

2. Rule Induction for Morphosyntactic Particle "De"

Using the basic ideas of rough set theory, learning from examples module version 2 (LEM 2) is adopted as the rule induction algorithm based on corpus with semantic tagging. As we known, LEM 2 is one of rule induction methods in LERS data mining system, the flow chart is illustrated in Figure 1.
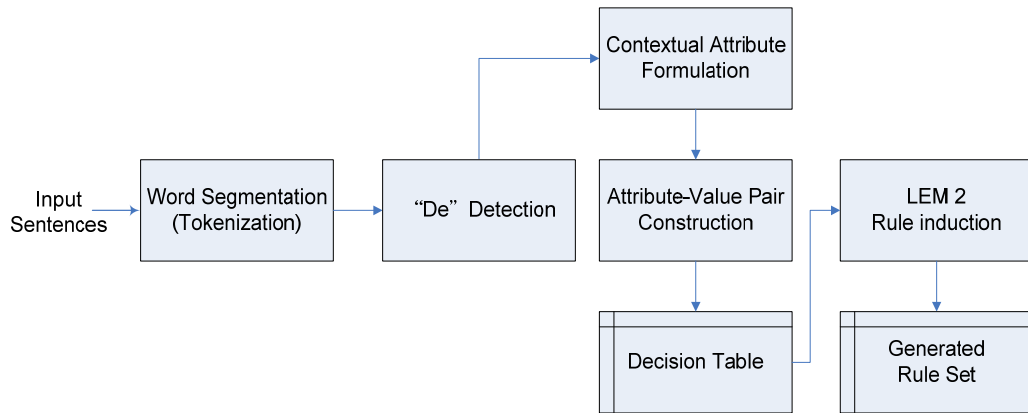

Figure 1. The LEM 2-based rule induction for the morphosyntatic particle "De."

This paper adopted the LEM 2 algorithm to natural language processing especially for Chinese information processing. For each input Chinese sentence, word segmentation is applied to obtain the word level tokens with part-of-speech (POS) tagging. The detection process for "De" is further used to select the sentence with "De." The sentences without "De" are dropped in the post-processing here. For extracting the linguistic feature to decide which morphosyntactic particle is used in the sentence, the contextual attributes are defines according to word and part-of-speech based n-grams. The contextual attributes accompanied with morphosyntactic particle {的, 得, 地} to constructing the attribute-value pairs. All the attribute-value pairs gathered in training data are fed into LEM 2 rule induction algorithm to generate the rule set. Therefore, the rule set can be further used to decide the usage of the morphosyntactic particle {的, 得, 地}. Herein, the proposed method is divided into two parts, decision table construction and LEM 2 rule induction algorithm, are described dentally in Section 2.1 and 2.2 respectively.

2.1 Decision Table Construction

Since the decision table is defined as a form for blocks of attribute-value pairs, the attribution plays an essential role in rule induction using LEM 2 algorithm. However, the sentence in natural language is not structural and fitting to the format of attribute. It is noteworthy how to transform the natural language into the attribute. That is to say, proposition extracted from sentence is one of the important issues for attribute. Herein, the contextual information surrounding the morphosyntactic particle "De" is used to form the attributes as shown in Figure 2.
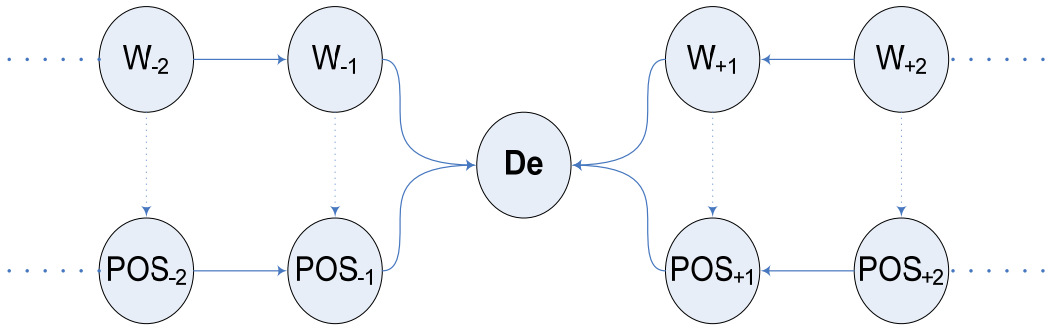

Figure 2. The contextual attribute formulation using the word with POS information surrounding the target word "De."

Each sentence representing one case and the independent variables are called attributes. As shown in Figure 2, the surrounding words with part-of-speech will combined with their relative position for particle "De" will combine into considering to form the attributes. The values are defined as one of the single-character words {的, 得, 地} in particle set. Each attribute-value pair represents one sample of knowledge about a decision table or a property of cases. These attribute-value pairs and the corresponding blocks serve as a basis for rule induction. Similar to N-gram models, the utility of the proposed contextual features is closely linked with the observation window size. As we know, the longer word sequence can provide more information for predicting the next word in N-gram models. This phenomenon leads us to find the near optimal window size for the "De" classifier. However, we have observed the empirical results of the larger windows size. Here, the relative positions from -2 to +2 are included in the windows for obtaining the contextual attribute, because the performance is near to those by the larger windows size. This condition not to conform to our expectation and the reason should be the limitation of the training corpus. For the example shown in Figure 3, the related information in decision table is illustrated in Table 1. The sentence containing the word sequence "欣賞(enjoy) 美麗(beautiful) 的(De) 一幅(a) 畫(picture)" is illustrated as the case 1 shown in Figure 3. Basically, each case is obtained from one sentence. Actually, the number of cases is dependent on the number of the particle "De" in the sentence. An

example　"特別(special)　的(De1)　愛(Love)　給(give/for)　特別(special)　的(De2)　您(you)" with two particles, the cases 2 and 3 is obtained from the same sentence. The cases 4 and 5 in Table 1 show the examples for　"得(De)"　and　"地(De)"　separately.
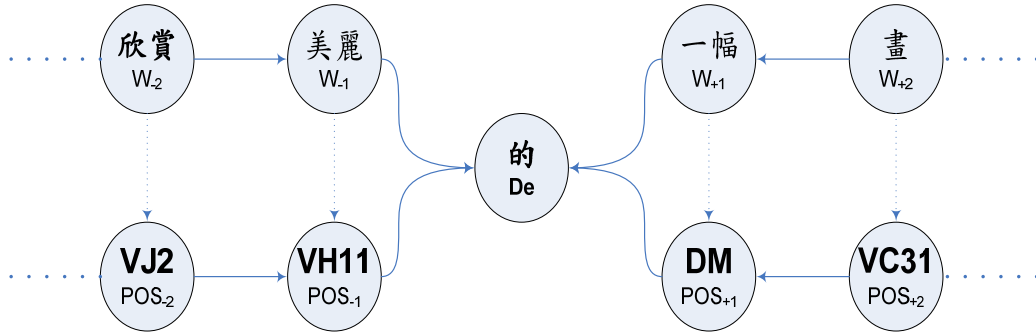

Figure 3. The contextual attribute formulation for the sentence containing
"欣賞(enjoy)　美麗(beautiful)　的(De)　一幅(a)　畫(picture)."

Table 1. A decision table for the decision of "De."

case	W$_{-2}$	POS$_{-2}$	W$_{-1}$	POS$_{-1}$	W$_{+1}$	POS$_{+1}$	W$_{+2}$	POS$_{+2}$	De
					Attributes				**Decision**
1	欣賞	VJ2	美麗	VH11	一幅	DM	畫	VC31	的
2	-	-	特別	VH11	愛	Nad	給	VD1	的
3	給	VD1	特別	VH11	您	Nhac	-	-	的
4	-	-	快樂	VH21	不得了	VH11	-	-	得
5	暗自	Dh	悲傷	VH21	低聲	Dh	哭泣	VA4	地
.				.					.
.				.					.
.				.					.

2.2 LEM 2 Rule Induction Algorithm

Rule induction is important to find relationships between blocks defined by condition attributes and the blocks defined by the decision attributes. In Chinese, particles usually connect adverb, adjective, verb and noun from the observations. *U* and *A* denote the set of all cases and the set of all attributes in decision table. Independent variables are treated as attributes. The target variable depended on attributes is called as decision. A function $f(\cdot)$ is defined for mapping the direct product of *U* and *A* into the set of all values. These

terminologies are defined in rough set theory and LEM 2. The fundamental idea of rough set theory is aimed at using the blocks in decision table to explain the rule induction. *Q* is one of the nonempty subsets of *A*. the indiscernibility relation *IND(Q)* is defined as follows.

$$(x, y) \in IND(Q) \Leftrightarrow f(x,a) = f(y,a) \ \ for \ \ \forall a \in Q \,, \tag{1}$$

Since the indiscernibility relation *IND(Q)* is an equivalence relation. The elementary set of *Q*, denoted by *[x]_Q*, is defined as the equivalence classes of *IND(Q)*. *IND(Q)* can be used to obtain the idea of blocks of attribute-value pairs. The intersections of blocks are shown in equation (2).

$$[x]_Q = \cap \{(a,v) \mid a \in Q, f(x,a) = v\} \,. \tag{2}$$

This investment adopted the rule induction algorithm to explore the search space of attribute-value pairs. Lower approximation for concept is defined as conditional probability is one. The probability of the upper approximation is greater than zero. According to lower and upper approximations, the concept is further divided into three areas: positive region, boundary region, and negative region. LEM 2 explores the search space of attribute-value pairs and finds a local covering and then converts it into a rule set. The algorithm is shown in (Grzymala-Busse, 2005).

3. EXPERIMENTAL RESULTS

For evaluating the proposed method, the LEM 2 algorithm is adopted for classifying the usage of particle "De" for the learners in Chinese as the second language. We first induce the rule set using the word and its corresponding part-of-speech information by the general training corpus. Furthermore, the evaluation set using the test corpus gathered from the non-native speakers for Chinese. The confuse matrix, precision rate, recall rate and F1 measures are applied for assessment of the proposed method. Here, we illustrate data preparation, evaluation metrics, experimental results and discussion in the following sections in detail.

Table 2. The rule set induced by the proposed method.

rule	W_-2	POS_-2	W_-1	POS_-1	W_+1	POS_+1	W_+2	POS_+2	De
1				VC		P	到		的
2				VC		VCL	到		的
3				VK		VH	不得了		的
4				VH		VH	不得了		的
5				V-		Dfa			的

6	V-	V-	(V_2)'	的
7	VH	Nv	努力	的
8	Dfa	V-		的
9	D	V-		的
10	VK	VJ		得
11	VK	VK		得
12	D	VJ		得
13	D	VK		得
14		Na		得
15		Neu		得
16		Nv		得
17		Nc		得
18	VH	VA		得
19	VC	VA		得
20	VA	VA		得
21	(沒)'	VJ		得
22	Nc-(好)	VC		得
23	(漂亮)'	VE		得
24	VH	D	就	得
25	VH	VL		地
26	V-	Dfa		地
27	VC	Nh		地
28	VC	Nv		地
29	VHC	Nh		地
30	VHC	Nv		地
31		Na		地
32		Neu		地

3.1 Data Preparation and Evaluation Metrics

Since the goal of this paper aims at the usage classification for morphosyntactic particle "De," two corpuses, CYCCDC (Yeh *et al*., 2014) and FinalTest_SubTask2 in shared-task on Chinese Grammatical Error Diagnosis (CGED) (Yu *et al*., 2014), are employed for training corpus and test corpus respectively. CYCCDC is a conversational dialogue corpus form daily life. The recorded speech is collected and annotated as text transcript. Considering of the learners' usage in real life and learning about the capabilities in listening, speaking, reading and writing, CYCCDC is used for building the rule set. The test file FinalTest_SubTask2 is provided for evaluating the Chinese grammatical error diagnosis. The sentence is gathered from the learner for Chinese as the second language. However, the number of the sentence containing error usage about the morphosyntactic particle "De" is not large enough. The character in the morphosyntactic particle "De" character set {的,得,地} is randomly re-assigned as the character from the same character set to form our test corpus.

The goals of this approach are to detect whether an input sentence contained error usage of the morphosyntactic particle "De" and to identify the correct character/word. Table 3 shows a contingency table of the related hypothesis.

Table 3. Contingency table for the usage classification of the morphosyntactic particle "De."

Hypothesis		Condition		
		Positive	Negative	Total
Outcome	Positive	True Positive TP	False Positive FP	P
	Negative	False Negative FN	True Negative TN	N
	Total	TP+FN	FP+TN	P+N

There are three metrics were used to assessing the proposed method: precision rate, recall rate and F1 measure, they are formulated as equations (3), (4) and (5) separately.

$$Precision\ rate = \frac{TP}{P}, \tag{3}$$

$$Recall\ rate = \frac{TP}{TP+FN}, \tag{4}$$

$$F1\ measure = \frac{2 \times Precision\ rate \times Recall\ rate}{Precision\ rate + Recall\ rate}. \tag{5}$$

3.2 Experimental Results and Discussion

Tables 4 and 5 show the evaluation results of confusion matrices of the usage classification about the morphosyntactic particle "De" in frequency count and percentage separately. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. This part aims at finding the confusion status among the words {的, 得, 地} by the rule set induced from LEM 2 algorithm. From the observation about confusion matrices, the correction rate of "的" is the highest compared to the "得" and "地." Due to the occupation ratio of "的" is higher than the other two particles in training corpus and test set, the miss rate about "的" is less than 10 percentages. This is excellent output for practice use. However, there are many false alarm errors about "的" cause the accuracies of "得" and "地" is not good enough. Many induction rules resulted from the training cases in corpus focus on the "的". This condition makes the false alrm and reduce the precision of the other particles "得" and "地.

Table 4. Count-based Confusion matrix for of the morphosyntactic particle "De."

	的	得	地
的	1560	69	49
得	67	46	7
地	45	6	36

Table 5. Correction percentage confusion matrix for of the morphosyntactic particle "De."

	的	得	地
的	0.929781	0.041120	0.029201
得	0.558333	0.383333	0.058333
地	0.517241	0.068965	0.413793

Tables 6 illustrates the performance measure about morphosyntactic particle "De" including the metrics precision rate, recall rate and F1 measure. From this result, we can find that the performance of "的" achieve the best performance compared to those of "得" and "地." This is affected by the occupation ratio of particle significantly. Besides, According to the observation of the outcome data, we find that the characters '的,' '得' and '地' maybe part of the word with multiple characters such as "目的," "得意" and "土地". This condition cause the performance dramatically reduced. These errors usually come from the wrong word segmentation and the characters.

Table 6. The performance measure of the proposed method using precision rate, recall rate and F1 measure.

	Precision rate	*Recall rate*	*F1*
的	0.9459 (70/74)	0.5932 (70/118)	0.7292
得	0.4242 (28/66)	0.3590 (28/78)	0.3889
地	0.5686 (29/51)	0.4915 (29/59)	0.5273

4. CONCUSIONS

In this paper, we focus on rule induction on the usage of morphosyntactic particle "De" for the Chinese as the second language learners. The attributes that were formed from the surrounding words and the corresponding part-of-speech are adopted for attribute-value pairs. The training data is fed into the rule induction process. Here, LEM 2 algorithm is adopted here for deriving the rule set to classify {的,得,地} in this investment. The main contribution of this paper aims at the attribute-value pair formulation from the sentence in natural language. Considering of the contextual information, the position and part-of-speech of the surrounding words are used to form the independent variables. More than thirty rules are induced by LEM 2 algorithm. According to the observation about experimental results, we found the proposed method is workable and its performance is good enough in practice. We illustrate the confusion matrix and performance measure based on precision and recall rates. By this approach, the Chinese as second language learners can obtains the desired help in the usage of morphosyntactic particle "De".

REFERENCES

Bai, X., Liang, F., Blythe, H. I., Zang, C., Yan, G., & Liversedge, S. P. (2013). Interword spacing effects on the acquisition of new vocabulary for readers of Chinese as a second language. *Journal of Research in Reading*, *36*(S1), S4-S17.

Chang, L. Y., Xu, Y., Perfetti, C. A., Zhang, J., & Chen, H. C. (2014). Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language. *International Journal of Disability, Development and Education*, *61*(3), 288-305.

Chang, R. Y., Wu, C. H., & Prasetyo, P. K. (2012). Error diagnosis of Chinese sentences using inductive learning algorithm and decomposition-based testing mechanism. *ACM Transactions on Asian Language Information Processing (TALIP)*, *11*(1), 3.

Chen, H. Y., & Liu, K. Y. (2008). Web-based synchronized multimedia lecture system design for teaching/learning Chinese as second language. *Computers & Education*, *50*(3), 693-702.

Chen, H. C., Hsu, C. C., Chang, L. Y., Lin, Y. C., Chang, K. E., & Sung, Y. T. (2013). Using a radical-derived character e-learning platform to increase learner knowledge of Chinese characters. *Language Learning & Technology*, *17*(1), 89-106.

Chu, H. T., Tsai, W. S., & Lee, S. Y. (2014). Design of a Cloud Service for Learning Chinese Pronunciation. *Journal of Electronic Science and Technology*, 1, 011.

Chuang, H. Y., & Ku, H. Y. (2011). The effect of computer?based multimedia instruction with Chinese character recognition. *Educational Media International*, *48*(1), 27-41.

Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*(1), 86-114.

Grzymala-Busse, J. W. (2005). Rule induction. *Data Mining and Knowledge Discovery Handbook*, 277-294. Springer US.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269-279.

Hsiao, H. S., Chang, C. S., Chen, C. J., Wu, C. H., & Lin, C. Y. (2013). The influence of Chinese character handwriting diagnosis and remedial instruction system on learners of Chinese as a foreign language. *Computer Assisted Language Learning*, 1-19.

Hsiao, H. S., Chang, C. S., Lin, C. Y., Chen, B., Wu, C. H., & Lin, C. Y. (2015). The development and evaluation of listening and speaking diagnosis and remedial teaching system. *British Journal of Educational Technology*.

Huang, C. K., Hsin, C. O., & Chiu, C. H. (2010). Evaluating CSL/CFL website usability: A user-centered design approach. *Journal of Educational Multimedia and Hypermedia*, *19*(2), 177-210.

Jiang, Y., Wang, T., Lin, T., Wang, F., Cheng, W., Liu, X, Wang C. & Zhang, W. (2012, June). A rule based Chinese spelling and grammar detection system utility. *IEEE International Conference on System Science and Engineering (ICSSE) 2012*, 437-440).

Li, Y. H. A. (2012). de in Mandarin↔ e in Taiwanese. *Studies in Chinese Linguistics*, *33*(1), 17-40.

Lin, C. Y., Huang, C. K., & Chen, C. H. (2014). Barriers to the adoption of ICT in teaching Chinese as a foreign language in US universities. *ReCALL*, *26*(01), 100-116.

Looi, C. K., Chen, W., & Wen, Y. (2009). Exploring interactional moves in a CSCL environment for Chinese language learning. In *Proceedings of the 9th international conference on Computer supported collaborative learning*, Vol. 1, 350-359.

Lu, C. C., Hsieh, G., Chung, Y., Liu, C., Shih, C. W., & Hsu, W. L. (2014). Suggestions on curriculum content design in learning Chinese as a second language. *International Conference on Information Society (i-Society) 2014*, 339-340.

Paul, W. (2012). Why Chinese de is not like French de: A critical analysis of the predicational approach to nominal modification. *Studies in Chinese Linguistics*, *33*(3), 182-210.

Shi, Y., & Li, C. N. (2002). The establishment of the classifier system and the grammaticalization of the morphosyntactic particle de in Chinese. *Language sciences*, *24*(1), 1-15.

Simpson, A. 2000. On the status of modifying DE and the structure of the Chinese DP. In S.W. Tang and L.L. Chen-Sheng (eds.), *On the Formal Way to Chinese Languages*. Stanford: CSLI, 74-101.

Yang, Y. F. (2011). Engaging students in an online situated language learning environment. *Computer Assisted Language Learning*, *24*(2), 181-198.

Yeh, J. F., Lu, Y. Y., and Tan Y. S. (2014). CYCCDC: A Chiayi Chinese Conversation Dialogue Corpus. In *Proceedings of the 22nd International Conference on Computers in Education, workshop proceedings*, 7-12.

Yip, P.-C., & Rimmington, D. (2004). *Chinese: A comprehensive grammar*. New York: Routledge.

Yu, L.-C., Lee, L.-H., and Chang, L.-P. (2014). Overview of grammatical error diagnosis for learning Chinese as a foreign language. In *Proceedings of the 1st Workshop on Natural Language Processing Techniques for Educational Applications*, 42-47.

Waltraud, P. A. U. L. (2012). *The insubordinate subordinator de in Mandarin Chinese: Second take*. The Attributive Particle in Chinese (tentative title) [Frontier in Linguistics Series](ed. Sze-Wing Tang), Beijing:Peking University Pres.