Effective Recognition and Statistic of Engagement Resources in Persian Discourses

W. Li
Luoyang University of Foreign Languages
Luoyang, China

M.Z. Yi
Luoyang University of Foreign Languages
Luoyang, China

W.Y. Liu
School of Foreign Language
Linyi University,
Linyi, China

Abstract—**Appraisal theory has been focused on by more and more scholars because of its powerful function of analyzing interpersonal meanings, which not only pays attention to the contents and structure of discourses, but attaches more importance to speakers' or writers' attitude, stance and emotional tendency as well. Engagement, as one system of appraisal theory, is concerned with the sources of attitudes and different Engagement resources imply different attitudes and ideas of speakers or writers. Therefore, through interpreting the texts of different genres by means of Engagement, we can reveal the purpose and emotional tendency hidden behind the language. The paper offers an effective way of recognizing and calculating Engagement resources in Persian discourses in order to help the researchers in that field improve the efficiency of analysis.**

Keywords-appraisal theory; engagement; automatic recognition and statistic; Persian discourses

I. INTRODUCTION

According to Halliday, the three tasks that languages should finish are ideational function, interpersonal function and textual function [1]. The interpersonal function can reflect the interpersonal relationship and the status, identity, motive and attitude of speakers or writers. As for the analysis of interpersonal meanings, systemic-functional linguistics applies the two systems of mood and modality. In Martin's opinion, there exists a defect just by means of the two systems, which is the ignorance of semantics of evaluation. Therefore, Martin has put forward Appraisal Theory to make up for the shortcoming. Martin and Rose mentioned that Appraisal Theory focuses on the appraisal resources in the discourses and these resources can help readers or audiences understand the feelings and thoughts of the writers or speakers and then social relations will be built [2].

In the author's view, when we do some emotion and semantic analysis，we should not be limited to the contents, structures or forms of texts, but also focus on how these contents come into being and explore the meanings hidden between the lines. Because the language is closely related to ideology and contents of the life, people have consciously and unconsciously selected and classified the words and means of expression which is to be used before speaking. That is to say, the language does not necessarily reflect the objective world, and to a certain extent will reconstruct the world. For example, if a policeman kills a hostage by mistake, different reports of the media can result in different attitudes of the public towards him. So, as long as we can reveal the attitudes and values, etc hidden behind the expressions, then we will obtain a more objective and more detailed result whether we do some emotion analysis or semantic mining.

Appraisal theory is an effective way to solve this problem, which is used to dig deep meanings through the text level. However, as far as the author knows, presently in the aspect of the application of Appraisal theory, the judgment, labelling and counting of the appraisal recourses are basically accomplished manually. Therefore, when it comes to a huge corpus, the processing efficiency will decline a lot. Appraisal theory consists of three systems, which are Attitude, Engagement and Graduation. The paper will mainly attempt to implement computer automatic recognizing and counting Engagement recourses, in order to improve the efficiency of Engagement analysis.

II. RELATED WORK

Appraisal theory originated in the 1990s and has been introduced into China in the beginning of the 21st century. In the aspects of both theoretical research and theoretical application, Appraisal theory is developing rapidly. Related monographs and articles are also growing continuously and the theory has been applied to various types of discourses, including public speech, news discourse, letter words, television program and literary works and so on. For example, some researchers provided useful suggestions for research and teaching by exploring how the undergraduates expressed and negotiated the appraisal meanings in the process of working through arguments [3]. Two scholars used Appraisal theory to analyze 9 articles on NATO's bombing of Serbia from the national post and the guardian in 1999 and found that the same topic may have different ways of displaying because of different standpoints of the authors [4]. Some scholars applied Appraisal theory to the sentiment analyzing and emotional processing of news texts and proved that Appraisal theory has strong ability in automated sentiment analysis of online news texts [5]. Another related paper discusses how Engagement analysis helps the authors align their audiences in time of

writing their grant proposals [6]. Still another one applies Appraisal Theory to wine tasting sheets [7].

To sum up, Appraisal theory provides a powerful tool for semantic mining and sentiment analysis.

III. ENGAGEMENT OF APPRAISAL THEORY

The theory explores the source of attitude. Different speakers have different standpoints, so interpersonal meaning negotiation and ways of communication and interaction with the audience or readers are not the same. Through analyzing Engagement recourses in the discourses, we can not only know the speakers' standpoint and emotional tendency, etc, but also find out how the speakers influence the audience's viewpoints by expressing opinion tendency implicitly.

Engagement is concerned with the diverse range of recourses by which speakers negotiate and adjust the arguability of their utterances. Engagement can be divided into "monoglossic" when the utterances make no reference to other voices and "heteroglossic" when they contain dialogic alternatives. Moreover, heteroglossic Engagement contains two domains: Dialogic Expansion and Dialogic Contraction. First, Table 1 shows the construction of Engagement [8] in macro view. Then we will illustrate all the subsystems briefly.

TABLE I. ENGAGEMENT SYSTEM.

Subsystem	Secondary Subsystem		Examples
Dialogic Contraction	Disclaim	counter	Yet, although, amazingly, but
		deny	no, didn't, never
	Proclaim	concur	naturally, obvious, of course, admittedly, sure
		pronounce	I contend, indeed, the facts of the matter are…
		endorse	The report shows/proves/demonstrates that…
Dialogic Expansion	Entertain		perhaps, probably, maybe, it seems to me, apparently
	Attribute		It's said that, the report states, Halliday argues that, many Australians believe…

Contractive resources refer to "meanings which construe a dialogistic backdrop for the text of other voices and other value positions, towards excluding certain dialogic alternatives from any subsequently communicative interaction or at least towards constraining the scope of these alternatives in the colloquy as it henceforth unfold". This subsystem consists of disclaim and proclaim. Disclaim is construed as "the textual voice which positions itself at odds with, or rejects some contrary position". To put it another way, we can think of it as "some dialogic alternatives which is directly rejected or supplanted, or is represented as not applying". It has two parts: counter and deny. Deny is "a resource for introducing the alternative positive position into the dialogue, and hence acknowledging it, so as to reject it". The deny resources are "seldom", "never", etc. Counter resources can be described as "formulations which represent the current proposition as replacing or supplanting, and thereby countering a proposition

which would have been expected in its place". For instance, "although", "amazingly", etc. Proclaim resources are "those formulations which, rather than directly reject or overrule a contrary position, act to limit the scope of dialogistic alternatives in the ongoing colloquy". Proclaim can be divided into endorse, pronounce and concur. Endorse refers to those resources by "which propositions sourced to external sources are construed by the authorial voice as correct, valid, undeniable or otherwise maximally warrantable". For example, "show", "proves", "point out", etc. Pronounce refers to that speakers make a point of view not easy to be questioned and against through obvious intervention or emphasis. For example, "in fact", "indeed", etc. Concur refers to that the speaker not only recognizes a point of view, but also expresses it as indisputable, such as "undoubtedly", "of course", "naturally", etc.

The purpose of expansion is to let other voices involved and thus open dialogue space. Expansion consists of entertain and attribute. Entertain resources are "those words by which the authorial voice indicates that its position is but one of a number of possible positions and thereby, to greater or lesser degrees, makes dialogic space for those possibilities". The speaker implies the view is not the only one and that other different views can also be involved, so the utterance space will be opened. For instance, "perhaps", "may", "in my opinion", etc. Attribute resources are "those formulations which disassociate the proposition from the text's internal authorial voice by attributing it to some external proposition from the text's internal authorial voice by attributing it to some external voices. This is most typically achieved through the grammar of directly and indirectly reported speech and thought". In other words, the speaker hides his own opinion in others' words and expresses his view by means of others' utterances. Of course, the speaker also does not exclude other different points of view in the meantime. The purpose is no need for being responsible for words. For example, "according to", "It's said that", "someone argues that", etc.

IV. METHODOLOGY

From the foregoing, Engagement resources can be divided into two categories, verbs and non-verb words. As for English and Persian, it is easy to find the corresponding words of the resources between the two languages. Appraisal theory as originated from the English language, so its application in English discourses has developed very well. At present, a lot of theses about the application of Appraisal theory are available and those theses usually have the part of corpus annotation in appendix. Through searching for the Engagement resources in 50 English discourses, we have found that the variety is quite limited. Maybe it results from the definition of Engagement. Then, we translate English resources into Persian ones by means of an online dictionary named Farsidic [9]. Farsidic is an online bilingual dictionary. There are two reasons for choosing it. On one hand, it is free and easy to use and on the other it is a general dictionary and it tends to give many translation options of the same meaning, which contributes to obtain as many Persian Engagement resources as possible.

The verb English Engagement resources are generally not infinitives, while a dictionary usually does not include the inflected forms of a word. Hence we need do some normalization for English words before translation. Then the Persian Engagement resources got by translation still need some processing. Firstly, if the ideographic part of the compound verb shows up alone, there are many meanings. So we need to add auxiliary verbs in at the time of record to prescribe a limit to the meaning of the compound verb. For example, the word نشان has many meanings. If we add دادن to it, its meaning will be limited to one. Because changes of the compound verb reflect on دادن, we need record all of its possible forms. Secondly, the ideographic part of some compound verbs is an adjective and it will express the meaning of a verb only together with auxiliary verbs. Such words also need to be recorded thoroughly in our database (eg. كرده فاش) . Thirdly, Persian also has a variety of writing standards, that is to say, the same word or phrase has more than one form (eg. تا به حال&تابحال , حتماً&حتما) . Fourthly, the Unicode of some Persian letters are fuzzy, so some letters likeك&ک, ي&ى have two forms. The former is the Persian Unicode and the latter is the Arabic Unicode. Finally, there are two kinds of spaces in Persian. They are white spaces (like spaces in English) and short spaces. They can appear in a word or between two words and at times no space is also OK (eg. فاش &فاش میکرد ,به طور تعجب‌آور&به طورتعجب‌آور ,بی تردید&بی‌تردید می کرد). All the possible cases above have been collected to the best of our ability.

This paper uses Access to store resources. As is shown in Fig. 1, there are Engagement resources in the left column and the right one represents categories.

FIGURE I.   PART OF THE DATABASE OF ENGAGEMENT RESOURCES

Then, we use C# to implement following functions. The first one is reading the Engagement resources in our database automatically. The second is retrieving the resources in the target documents in order to obtain the number of occurrence and proportion of all kinds of Engagement resources and listing the sentences owning the resources out. As is shown in Fig. 2,

FIGURE II.   INTERFACE OF THE PROGRAM

V.   EXPERIMENT

To study the effect of applying NLP techniques on the efficiency of recognizing and counting Engagement resources, we take Mr. Ahmadinejad's speeches [10] in the United Nations General Assembly as research corpora and his speeches have been annotated manually first. Then we calculate results via the software Wordsmith. Next the effect of applying NLP techniques will be evaluated by comparing the result obtained by Wordsmith and that of our program.

As for manual calculation, frequencies of Engagement resources are calculated via the function of Wordlist in Wordsmith. As the software only recognizes txt files, we put Persian speeches in text and store them in the coding of Unicode so as not to cause digital gibberish. Then a wordlist will be formed. It is convenient for us to count the English abbreviation (representing the category of Engagement resources) if we click the alphabetical button below the dialog box.

FIGURE III. SCREENSHOT OF THE WORDLIST IN WORDSMITH

Freq in the Wordlist stands for the occurrence number of the words, so the total number of Engagement resources

$$=(N_{att}+N_{ent})+(N_{den}+N_{coun}+N_{conc}+N_{pron}+N_{end})$$

$$=(17+43)+(14+34+12+20+3)$$

$$=143.$$

Then the following tables can be obtained.

TABLE II.   STATISTIC EFFECTIVENESS OF DIALOGIC CONTRACTION & DIALOGIC EXPANSION MANUALLY AND AUTOMATICALLY

Category	Percentage Obtained Manually	Percentage Obtained Automatically
Dialogic Contraction	58%	55%
Dialogic Expansion	42%	45%
Attribute	28%	30%
Entertain	72%	70%
Disclaim	58%	54%
Proclaim	42%	46%
Attribute	12%	13%
Entertain	30%	32%
Deny	9%	6%
Counter	24%	25%
Concur	9%	10%
Pronounce	14%	11%
Endorse	2%	3%

Through comparing the statistical results of two methods, we have found the percentages obtained manually and automatically are very similar. That is to say, our automatic method satisfies the need for Engagement system analysis.

VI. CONCLUSION

Because of its powerful analytical ability of attitude tendency, the Engagement system of Appraisal theory is widely used in various types of discourses. However, as the author knows, presently recognition and statistic of the resources in the process of analysis are realized manually, which is obviously powerless for the large corpora. Therefore, this paper puts forward a method that can recognize and calculate Engagement resources in discourses, which greatly improves the analytical efficiency of Engagement system and removes the limitations in amounts of the corpus. In addition, the paper combines Engagement system of Functional linguistics with Computational linguistics, also offering a train of thought for the sentiment analysis and semantic mining in Computational linguistics.

REFERENCES

[1] Halliday, M.A.K., An Introduction to Functional Grammar. London: Edward Arnold, 1994.

[2] Martin, J. R. & David, R., Working with Discourse: Meaning beyond the Clause. Beijing: Peking University Press, 2003/2007.

[3] Wu, S. M. & Allison, D., Exploring appraisal in claims of student writers in argumentative essays. *Prospect*, 23 (5), pp. 59-75, 2003.

[4] Jovanovich-Krsic, V., Evaluating the discourse of war in the press media: a lexico-grammatical examination of the 1999 NATO bombing of Serbia from the perspective of Appraisal theory. *Dissertation Abstracts International*, 66 (1), page: 0188, 2004.

[5] Khoo, C. S., Nourbakhsh, A. & Na, J.-C., Sentiment analysis of online news text: a case study of Appraisal theory. *Online Information Review*, 36 (6), pp. 858-878, 2012.

[6] Pascual, M. & Unger, L., Appraisal in the research genres: An analysis of grant proposals by Argentinean researchers. *Revista Signos*, 43(73), pp. 261-280, 2010.

[7] Breit, B. W., Appraisal Theory applied to the wine tasting sheet in English and Spanish. *Ibérica*, pp. 97-120, 2014.

[8] Martin, J. R. & White, P. R. R. The Language of Evaluation: Appraisal in English. Beijing: Foreign Language Teaching and Research Press, 2005/2008.

[9] Farsidic, http://www.farsidic.com.

[10] Available at: http://www.president.ir.