

Using Grounded Theory and Text Mining to Find Interesting Reading Materials for Slow EFL Learners

Yuh-Chang Lin¹, Chia-Ling Hsu², Mu-Hua Lin³, Hsiao-Fang Yang³,
and Chao-Fu Hong⁴

¹ Instructor, Center for General Education, Aletheia University,
No. 32, Chen-Li Street, Tamsui,
New Taipei City 25103, Taiwan, R.O.C.

² Graduate Institute of C&I, Tamkang University,
No. 151, Ying-chuan Road, Tamsui,
New Taipei City, 25137, Taiwan, R.O.C.

³ Department of MIS, National Chengchi University,
No. 64, Sec. 2, Chihnan Rd.,
Mucha Dist, Taipei 11605, Taiwan, R.O.C.

⁴ Department of Information Management, Aletheia University,
No. 32, Chen-Li Street, Tamsui,
New Taipei City 25103, Taiwan, R.O.C.
au1258@mail.au.edu.tw, clhsu@mail.tku.edu.tw,
{kikalina, hfyang.wang}@gmail.com,
au4076@au.edu.tw

Abstract. Many reasons contribute to slow EFL learners. Of all the reasons which cause slow EFL learners, reading materials not appealing to learners tend to be the one to blame. In order to help teachers to find out interesting reading materials for slow EFL learners in Taiwan, this study intends to use Grounded Theory [8] and Text Mining to search for the reading materials which are appealing to slow EFL learners in Taiwan. Based on the approaches of free voluntary reading proposed by Krashen [10], and extensive reading advocated by Day and Bamford [7], the theoretical framework of this study is established. Krashen claimed that when students read for information, pleasure or problem-solving, they have acquired the essence of free voluntary reading. Day and Bamford [7] argued that extensive reading is an important way to reinforce the concept of learning to read by reading. Both of these two reading approaches put emphasis on learners' free choice of reading materials which are within their linguistic proficiency, and at the same time, are interesting to them. Eighty-three university students are the participants, who took the second-year English reading course in a summer session program in Taiwan. They failed the course in regular semesters. This is their second- or third-time taking the same course. We call them slow learners accordingly. By the employment of Grounded Theory and Text Mining, it is expected that the interesting reading materials can be discovered and provided for teachers to adopt in their instruction in classes, and to a certain extent, learners' attention to reading passages can be drawn.

Keywords: Grounded Theory, Text Mining, interesting reading materials, EFL learners, free voluntary reading, extensive reading.

1 Introduction

1.1 The Importance of Reading

Reading plays an important role in literacy education and many research results have proven this. Krashen [10] advocated that his free voluntary reading is a way to develop one's literacy as well as to read for information, pleasure, and general knowledge. Day and Bamford [7] proposed that extensive reading is an approach to literacy education and only by extensive can learners acquire knowledge and information subconsciously or even unconsciously. Both of these two reading approaches share a common ground: Readers can choose what they like to read and read abundantly. For first language literacy education, free voluntary reading and extensive reading are widely employed to achieve the goal. However, it is not the case for reading in EFL contexts. This study aims to find out interesting reading materials for EFL learners with a view to attracting their attention to reading. Eventually, they may read extensively and at the same time they may develop their linguistic competence as well as their knowledge of the world.

1.2 Current Status of English Education in Taiwan

Currently, most studies on EFL reading in Taiwan, especially in the stage of primary and secondary English education, tend to focus on how to improve students' linguistic competence. It is this purpose that many teachers as well as materials designers make their efforts to use or produce materials that focus simply on the form of a language rather than the meaning of it. In other words, much emphasis is put on the grammatical aspects of a language: recognizing parts of speech for a word, distinguishing SVO grammatical pattern (subject-verb-object sequence of an English sentence), and acquiring phonetic system of English. Most of the efforts made both by teachers, materials designers, and students are for the purpose of mastering the form of English. Meaning of the content is rarely emphasized. That is the reason why students cannot recall accurately what they have read when they are asked to talk about the content of the reading passages they have just read. Students did not expose themselves to reading sufficiently. What with the reading materials chosen by teachers are not within students' linguistic proficiency, and what with the reading materials selected by teachers are not appealing to students, lead to students' lack of reading, which impedes students' significant improvement in English language ability.

1.3 The Significance of Meaning

Acquiring meaning of a reading passage will gradually build up a person's cognition process. From the accumulation of one's cognition of the reading materials or any events happening around, one may establish his or her schemata by turning facts into knowledge. When a person has much knowledge of the world, plus required linguistic ability, he or she will read both accurately and fluently. Meaning here plays an

important role in the formation of one's schemata or background knowledge. It is meaning rather than form of a language which is responsible for the acquisition of a language. Free voluntary reading and extensive reading are the efficient approaches to equipping learners with as much general knowledge as possible. However, how can we make these two approaches possible in a foreign language context is another issue? Going back to the shared ground for both free voluntary reading and extensive reading, we may find two principles: One is that readers can choose what they like to read, and the other is that readers read abundantly. This is the theoretical base for this study: Finding out the materials interesting to students in order that both free voluntary reading as well as extensive reading can be made possible. It is hoped that through this study, materials appealing to learners can be found and learners' attention can be drawn because they are interested in the materials, for the attention readers pay to particular elements of the text is associated with interest [12][13][14].

1.4 Situational and Individual Interest

Interest can be divided into two alternative characterizations: situational and individual or personal interest [1]. Situational interest is centered in the immediate environment and is typically regarded as transitory or fleeting [4]. Because situational interest entails getting learners' attention and keep them excited or enthused, it can be a positive influence in students' text-based learning. As to individual or personal interest, it is enduring and reaches into an individual's cognitive and affective nature [2][4]. Individuals' vocations and avocations are indicative of their personal interests [11]. Moreover, these deep-rooted interests are strongly associated with self-concepts and self-schemata [3]. It is the second type of interest that this study is searching for.

1.5 Grounded Theory

Grounded Theory comprises a systematic, inductive, and comparative approach for conducting inquiry for the purpose of constructing theory [5][6]. It can be divided roughly into three main stages: open coding, axial coding, and selective coding. In the stage of open coding, researchers are fracturing and analyzing the collected data. The core category or related concepts tend to emerge from the data collected in this stage. And then through theoretical sampling and selective coding of data and constant comparison of incidents or indicators in the data to elicit the properties and dimensions of each category, the theoretical saturation of the core and related concepts tends to be achieved. It is this constant comparing of incidents that an interchangeability of indicator can be obtained, which means that no new properties or dimensions are emerging from continued coding and comparison. At this point, since the theoretical situation of the concepts has been achieved, the researchers shift the attention to the emergent fit of potential codes that make the conceptual integration of the core and related concepts to produce hypotheses that help explain relationships between concepts, which accordingly accounts for the latent pattern of social behavior that form the basis of the emergent theory.

1.6 Text Mining

Text mining is a process to acquire high-quality information from text. High quality in text mining usually refers to some combination of relevance, novelty, and interestingness, which is typically derived through the divining of patterns and trends by means of statistical pattern learning. Text mining involves the process of structuring the input text, which refers to paring, along with the addition of certain acquired linguistic features and the removal of others, and subsequent insertion into a database. Typical text mining tasks include text categorization, text clustering, concept or entity extraction, sentiment analysis, document summarization, and entity relation modeling.

1.7 Slow EFL Learners

Participants are 83 university students who took second-year English reading course in the summer session. This is a make-up class for students who have failed in regular semesters. Most of them took this course for second time, and part of them even takes the course for more than once. This is the reason why we call them slow learners. Their linguistic competence is not good, nor is their general knowledge of the world. Because of the lack of their linguistic competence, they find it hard for them to read in English. Grabe [9] argued that readers will employ the schemata or background knowledge acquired in the first language to read in a second or foreign language when their “level of second language (L2) proficiency has been developed first so that first language (L1) academic language abilities can more readily transfer and have an impact on L2 reading (pp. 145-146)”. Interest and motivation are two other factors which will influence readers’ willingness to reading. If readers are interested in certain topics, they tend to read them regardless of their poor language ability, for interest helps ease their difficulty with the language and makes extensive reading possible. In other words, with interest, readers may read for pleasure, information, and solving problems [10].

2 Procedure

Eighty three students are asked to write out their idea on two questions: Do you like to read? If you do, what do you like to read? Students answered these two questions in Chinese for around three paragraphs within an hour. After they have finished writing up their answers to these two questions, the data were collected and typed accordingly. Then open coding was conducted. In this stage, the researcher was fracturing and analyzing the collected data. First of all, terms related to topics and genres which appeal to learners are kept and conceived of as meaningful concepts. For example, terms such as novels, magazines and comic books are kept and taken as subcategories of genres. Additionally, reasons for the preference for certain types of genres or topics are also kept in order that new, interesting and novel scenario can be discovered. After this process, axial coding has been administered to find out new, interesting, and novel storylines which are embedded in the data collected. Finally,

selective coding was conducted to find a scenario which is interesting and novel for the whole data collected.

3 Findings

By means of the application of human machine interaction, via three-stage coding process derived from the grounded theory, and through the employment of VFT Textmining System, we obtained four figures. They are shown as follows:

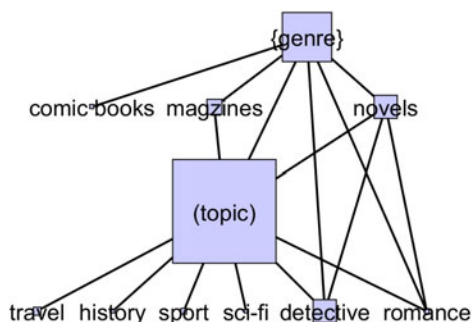


Fig. 1. A whole view of the extracted data

From the first figure, we may clearly observe that novels, magazines, and comic books are the main three categories which appeal to slow EFL university learners. These three categories can be referred to as the subcategories of genres. In other words, genre is the superordinate of novels, magazines and comic books. Also, we may see that novels have two subcategories: romance and detective. As to the subcategories of science fiction, sport, history, and travel, we are not sure whether they belong to the categories of novels or magazines. We cannot figure them out until we go on to observe the second figure.

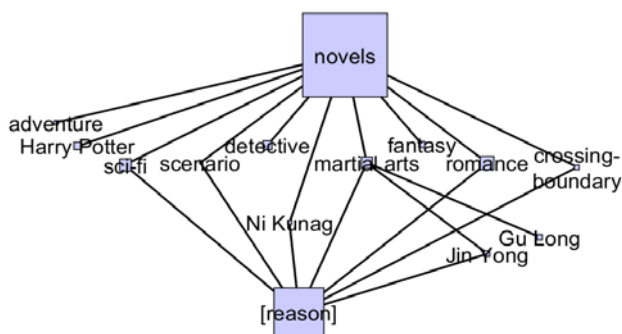


Fig. 2. Subcategories of novels

From the second figure, we may see that novels include such subcategories as Harry Potter, detectives, adventures, science fictions, fantasies, martial arts, romance, and crossing boundary. From this figure, we may find that science fiction belongs to the category of novels, and the rest of three subcategories (travel, history, sport) found in the first figure remains unidentified. Before we move on to find out the result, we want to point out two interesting discoveries. One is that students like to read Harry Potter; the other is that such Chinese authors of martial arts as Gu Long and Jin Yong are popular with students. Another Chinese author of science fiction, Ni Kuang was also admired by students. There is another finding that some students like to read so-called ‘crossing-boundary novels’, which is actually a kind of Internet novel mixed with the elements of martial arts, romance and fantasy. By utilizing this figure, we may come to a conclusion that the subcategories of science fiction and history are connected to the category of novels. Obviously, the category of magazine covers the subcategories of sport and travel.

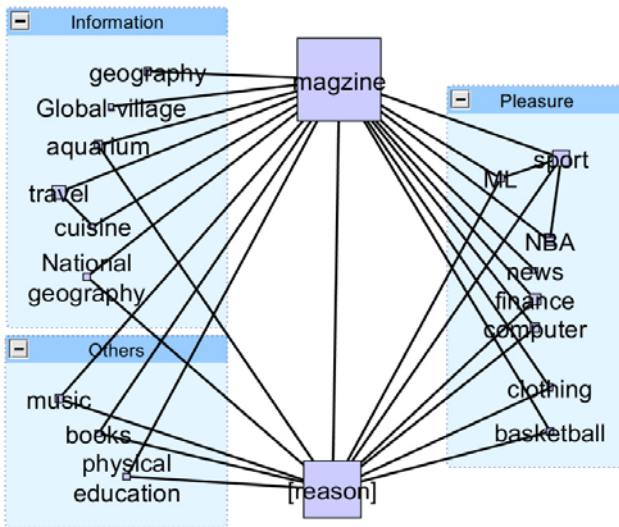


Fig. 3. Subcategories of magazines

In the third figure, we may find that below the category magazine, there are various kinds of subcategories such as newspaper, finance, computer, clothing, aquarium, basketball, Global Village (a kind of language learning magazine in Taiwan), travel, cuisine, National Geography, National Basketball Association, sport, physical education, music, and Chinese Professional Basketball. Of them, sport and travel are two main subcategories. We can roughly divide these subcategories into three items: information, pleasure, and others. Information covers items such as geography, global village, aquarium, travel, cuisine, National Geography. The items of pleasure include sport, major league, National Basketball Association, news, finance, computer, clothing, and basketball. Others contain such items as music, books, physical education.

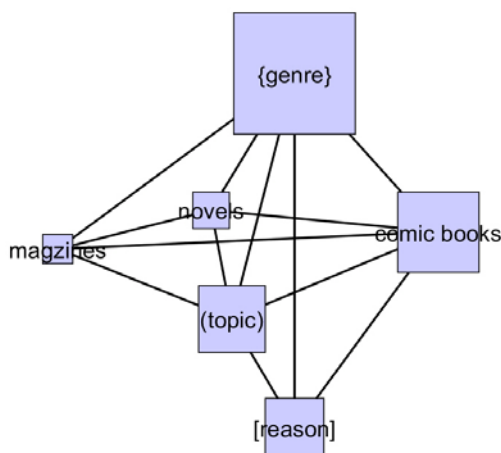


Fig. 4. An association map built based on the main category of comic books

There is no subcategory derived from the treated data, since students just mentioned comic books and the subcategories were not further elicited.

4 Discussion

As far as genre is concerned, novels, magazines, and comic books are the three main categories which appeal to students. When we take close look at these categories, we may find that in the category of novels, students like to read something about romance, detective, sci-fi, and history. In the category of magazines, students prefer sport and travel. For students who like to read novels, we may find a phenomenon that they like to read them both in traditional written form (namely books) and in the virtual world. Harry Potter and martial arts series are the ones in the traditional written form, and the so-called “boundary-crossing novels” are the combination of martial arts, romance, and fantasy, which are shown in the virtual world, that is, on the Internet world. This is a reflection of the computer generation, in which virtual worlds are created for people to fulfill some of their dreams which they fail to achieve in the real world.

Of all the subcategories connected to the main three categories (novels, magazines, and comic books), crossing boundary is the one which worth noting for the following two reasons. Firstly, it is at least new to me, even though I have been teaching English for almost 23 years. Secondly, crossing-boundary is a byproduct of the computer generation for it is a combination of romance, martial arts, and fantasies, which is presented in the virtual world. We may refer this phenomenon to the easy accessibility to computers. People increasingly count on computers for the management of their daily routines. They may read on the Internet for a variety of materials. Crossing-boundary novels in the virtual world are the reading materials appealing to university students. As teachers, it is one of their jobs to understand their students’ interest so that they may provide their students with reading materials they are interested in.

From both categories and subcategories derived from this study, we may come to a conclusion that learners are interested in reading for either information or pleasure. For example, in figure 3, we may find that the subcategories of magazines can be roughly divided into three items: information (including geography, Global Village, aquarium, travel, cuisine, and National Geography), pleasure (including sport, ML, NBA, news, finance, computer, clothing, basketball, and others (including music, books, physical education). In the first sight, learners just read for information and pleasure, how can they improve their language ability? The answer is definitely confirmed by Krashen [10], for he argues that reading makes learners not only acquire the general knowledge of the world but also learn the language subconsciously.

Furthermore, from the following frequency table which is derived from the third figure, we may see the frequency of the items under the information category. The frequency for travel is 9, cuisine 5, aquarium 4, national geography 3, global village 2, and geography 2. When choosing teaching materials for learners to read, we should take the frequency of the items into consideration. Especially, when all the topics cannot be adopted in a time, we should cover the one or ones whose frequency is or are higher than others.

Table 1. Frequency table of the term under the category of information

Term	Frequency
travel	9
cuisine	5
aquarium	4
national geography	3
global village	2
geography	2

5 Limitation of the Study

The data collected in this study was limited, for the 83 students were from a class in a university in Taiwan. Even though they were slow EFL learners as stated above, they were actually not the whole population of a kind in Taiwan. However, the data analyzed are still worth paying attention to, for they stand at least for certain portion of learners who express themselves directly concerning their preference for the topics and genres when they read. Of course, more extensive and multi-dimensional sampling is needed in order that the theoretical saturation can be achieved.

6 Further Study

Finding out the interesting topics or genres for those slow learners is the first step to get slow EFL learners involved in the learning of a second language. However, it is not sufficient for language teachers to simply provide their students with the materials which appeal to their students. As is the case when teachers provide students with the topics or genres which students are interested in, students still find themselves unwilling to or incapable of reading the materials provided by the teachers because of

too many unfamiliar words, phrases or sentence patterns they encounter while they are reading. It is a further study for researchers to search for materials which are within students' language ability, especially within students' word power. When learners read what they like to read and read without frequently stopping to look up unfamiliar words, they may read fluently. When they read fluently, their interest in the activities of reading may be stimulated. When students are interested in reading, they may read abundantly. When students read abundantly, the essence of extensive reading advocated by Bamford, Day and Krashen can be achieved. When students read extensively, the ultimate goal of reading for pleasure, information, and general knowledge [10] can be made possible.

References

1. Alexander, P.A., Jetton, T., Kulikowich, J.M.: The interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology* 87(4), 559–575 (1995)
2. Alexander, P.A., Kulikowich, J.M., Schulze, S.K.: How subject-matter knowledge affects recall and interest on the comprehension of scientific exposition. *American Educational Research Journal* 31, 313–337 (1994)
3. Alexander, P.A.: Knowledge seeking and self-schema: A case for the motivational dimensions of exposition. *Educational Psychologist* 32, 83–94 (1997a)
4. Alexander, P.A.: Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In: Pintrich, P.R., Maehr, M.L. (eds.) *Advances in motivation and achievement*, vol. 10, pp. 150–213. JAI Press, Greenwich (1997b)
5. Charmaz, K., Henwood, K.: Grounded theory in psychology. In: Willing, C., Station-Rogers, W. (eds.) *Handbook of Qualitative Research in Psychology*. SAGE, London (2007)
6. Charmaz, K.: *Constructing Grounded Theory: A practical guide through qualitative analysis*. SAGE, London (2006)
7. Day, R., Bamford, J.: *Extensive Reading in the Second Language Classroom*. Cambridge University Press, New York (1988)
8. Glaser, B.G., Strauss, A.L.: *The discovery of grounded theory: Strategies for Qualitative Research*, Chicago, Aldine (1967)
9. Grabe, W.: *Reading in a Second Language: Moving from Theory to Practice*. Cambridge University Press, New York (2009)
10. Krashen, S.D.: *The Power of Reading*, 2nd edn. Heinemann, Portsmouth (2004)
11. Mitchell, M.: Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology* 85, 424–436 (1993)
12. Reynolds, R.E., Shirey, L.L.: The role of attention in studying and learning. In: Weinstein, C.E., Goetz, E.T., Alexander, P.A. (eds.) *Learning and Study Strategies: Issues in Assessment, Instruction, and Evaluation*, pp. 77–100. Academic Press, San Diego (1988)
13. Reynolds, R.E.: Learning important information from text: The role of selective attention. *Review of Educational Psychology* 4, 345–391 (1992)
14. Shirey, L.L.: Importance, interest, and selective attention. In: Renninger, K.A., Hidi, S., Krapp, A. (eds.) *The role of interest in learning and development*, pp. 281–296. Lawrence Erlbaum Associates, Hillsdale (1992)