ELSEVIER

A critical evaluation of text difficulty development in ELT textbook series: A corpus-based approach using variability neighbor clustering

CrossMark

Alvin Cheng-Hsien Chen

Department of English, National Changhua University of Education, Changhua City, 500, Taiwan

ARTICLE INFO

ABSTRACT

Although the importance of English Language Teaching (ELT) textbooks is widely acknowledged, previous evaluation of ELT materials has paid little attention to the appropriateness of the text difficulty development in a textbook series. The present study aims to assess the progression of text difficulty in different textbook series in Taiwan, the rationale of which is argued to be generalizable to other ELT contexts. Specifically, there are two methodological emphases. First, text difficulty has been quantitatively measured by the BNC corpus-based frequency lists and a comprehensive set of well-established readability formulas, considering both vocabulary and structure complexity of the texts; second, a clustering-based statistical algorithm—variability neighbor clustering—is utilized to identify the developmental stages in text difficulty on an empirical basis. This corpus-based computational method not only objectively determines the developmental gaps in a textbook series, but also identifies the direction of the difficulty progression in vocabulary and structure complexity. This rigorous textbook evaluation provides a common framework for the assessment of text difficulty progression in the ELT materials. Several pedagogical implications are drawn for EFL learners and teachers as well as ELT textbook developers.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Although the importance of English Language Teaching (ELT) textbooks is widely acknowledged, evaluations for ELT materials development are still not a "well supported project" (Ghorbani, 2011). Many scholars have suggested checklists for a thorough examination of the course book contents, considering a wide range of critical features, such as practical considerations, four-skill balance, exercises and activities, pedagogical analyses, appropriateness in language and grammar, and supplementary materials (Ghorbani, 2011; Mukundan & Ahour, 2010; Tomlinson, 2012; Tsagari & Sifakis, 2014). However, these checklists tend to be researcher-dependent and context-dependent, paying little attention to the appropriateness of the text difficulty development in a textbook series.

The purpose of this study is to address this issue in the context of the official English curriculum in Taiwan senior high school. After the full-scale deregulation of textbooks in 2001, many publishers in Taiwan began to edit and publish their own versions of course books, following the guidelines and wordlists regulated by the national government. One of the major

E-mail address: alvinworks@gmail.com.

criticisms from teachers is the inadequacy of the difficulty development in the current textbook series, in which volumes used by the higher graders are not necessarily more difficult than those used by the lower graders. This inconsistency often results in a compromise that teachers may select different versions of the textbooks for different academic years. This *ad hoc* decision may accidentally deprive learners of a chance for a systematic exposure to a grammar illustration provided by the publisher.

Previous research on textbook development is somewhat limited in scope, often focusing on a particular linguistic aspect of the materials rather than the overall appropriateness of the text difficulty development. For example, a series of studies has been dedicated to the appropriateness of idiomaticity in the ELT materials by comparing the vocabulary distribution (Hsu, 2009), collocation patterns (Tsai, 2015), phrasal verbs (Chuang & Tsai, 2009), multi-word sequences (Lin, 2014), or formulaic sequences (Hsu, 2014b) of the ELT textbooks in Taiwan with those of the native-speaker production in different representative corpora. These comparisons have taken the ELT materials as a whole in their assessment and failed to provide a critical evaluation for the text difficulty progression in different volumes of the textbook series. Furthermore, the diversity of the linguistic emphases in previous studies has also suggested that a principled framework for ELT materials evaluation is yet to come (cf. Tomlinson, 2012).

There have been two main lines of research in the literature on the assessment of text difficulty in high-school textbooks in Taiwan. One of the research lines was to provide a descriptive account of the vocabulary distribution in the textbooks across different frequency-based word lists (Chen, 2014; Kao, 2014; Ting, 2005). They were concerned with the lexical coverage of the text materials, i.e. the percentage of the words that a reader understands or a reference word list covers. While the vocabulary coverage of different textbook versions have been compared, little attention has been paid to the progression between volumes in one series. Moreover, the discussion in the previous studies was often limited to a descriptive comparison with respect to the designated frequency word bands (e.g. the percentage differences in the first 1000-word band, 2000-word band etc.). The variation across all relevant frequency word bands was not analyzed as a whole, thus failing to provide a holistic account of the progression of the overall vocabulary difficulty across different volumes/versions. Another line of research on high school course books focused on the changes of readability across different book levels in a series (Chiu, 2010; Lin, 2008; Lo, 2010; Yeh, 2003). Quite a range of readability formulas were adopted in different projects depending on the limitation and the availability of the analytic instruments and computer programs (e.g. Gunning Fog Index in Chiu (2010), Fry in Lin (2008), Lix in Yeh (2003), and Flesch Reading Ease in Lo (2010)). Yet different readability formulas make commitment to different structural parameters (e.g. the average number of characters per word, the average number of syllable per word, or the average number of words per sentence) in computing the grammatical difficulty of the text materials. It would be more comprehensive if the analysts could consider the variation in different readability formulas. Furthermore, the assessment of the difficulty progression in the previous research relies mostly on an inspection of the visual graph (i.e. a line plot of the readability index values).

The present study aims to bridge this gap by proposing an objective and quantitative method to evaluate the arrangement of text difficulty for a textbook series. Both vocabulary and structural complexities will be included in our assessment of text difficulty by incorporating corpus-based frequency lists and well-established readability formulas. Most importantly, a clustering-based statistical algorithm—variability-based neighbor clustering (VNC) (Gries & Hilpert, 2008)—is adopted to identify the developmental stages of text difficulty on a more empirical basis. It is argued that the rationales of our critical textbook assessment are generalizable to other ELT contexts. On the one hand, our operational definition for text difficulty provides a common ground on which the difficulty levels of ELT materials from different cultural communities can be easily compared with each other. On the other hand, the algorithm of VNC further implements a robust analysis on the progression of the text difficulty in a given textbook series, thus shedding light on the developmental gaps across different book levels in the series.

2. Measuring text difficulty

2.1. Vocabulary levels and reading comprehension

Comprehensive reading is one of the keys to successful language learning. EFL learners are expected to possess a critical mass of L2 knowledge, including inferring the meanings of the unknown words from context, identifying the argument structure, and distinguishing idiomatic constructions. Most importantly, learners' vocabulary knowledge plays a crucial role in their reading proficiency (Hu & Nation, 2000; Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Lin, Hue, Lin, & Hsu, 2003).

Quite a few studies have been devoted to investigating how many words an L2 reader needs to know (i.e. their vocabulary coverage) for an adequate reading comprehension (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1989, 1992; Laufer & Ravenhorst-Kalovski, 2010; Lorge & Chall, 1963). Before we discuss the results from these studies, two factors need to be more carefully considered with respect to the methodology. First of all, "adequate" reading comprehension is a tricky idea and its operational definition may vary from scholar to scholar. The confusion may result from (1) the types of evaluation an analyst adopts for reading comprehension; (2) the threshold an analyst defines as an adequate level of reading comprehension. On the one hand, a learner's reading comprehension of a given text can be evaluated by two types of assessments: reading tests or cloze tests. Reading tests refer to the typical comprehension-checking tests, as often seen in most standardized English proficiency tests such as GEPT, TOEFL, and TOEIC. The other type is a cloze test, first introduced by Taylor (1953), which uses a text with regularly deleted words (usually every fifth word) and requires the subjects to fill in the blanks. A score from a reading comprehension test may not necessarily be comparable to one from a cloze test. On the other