

# EFL Learner Reading Time Model for Evaluating Reading Proficiency

Katsunori Kotani<sup>1/2</sup>, Takehiko Yoshimi<sup>3/2</sup>, Takeshi Kutsumi<sup>4</sup>,  
Ichiko Sata<sup>4</sup>, and Hitoshi Isahara<sup>2</sup>

<sup>1</sup> Kansai Gaidai University  
16-1 Nakamiya Higashino-cho, Hirakata, Osaka, Japan  
kat@khn.nict.go.jp

<sup>2</sup> National Institute of Information and Communications Technology  
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan

<sup>3</sup> Ryukoku University  
1-5 Yokotani, Seta Oe-cho, Otsu, Shiga, Japan

<sup>4</sup> Sharp Corporation  
492 Minosho-cho, Yamatokoriyama, Nara, Japan

**Abstract.** We propose a reading time model for learners of English as a foreign language (EFL) that is based on a learner's reading proficiency and the linguistic properties of sentences. Reading proficiency here refers to a learner's reading score on the Test of English for International Communications (TOEIC), and the linguistic properties are the lexical, syntactic and discourse complexities of a sentence. We used natural language processing technology to automatically extract these linguistic properties, and developed a model using multiple regression analysis as a learning algorithm in combining the learner's proficiency and linguistic properties. Experimental results showed that our reading time model predicted sentence-reading time with a 22.9% error rate, which is lower than the models constructed based on linguistic properties proposed in previous studies.

## 1 Introduction

One of the critical issues in learning or teaching a foreign language is learners' individual differences in proficiency. Unlike first language acquisition, proficiencies in acquiring a foreign language vary greatly. Thus, a language teacher has to understand each learner's problems and help the learner contend with them. The learners' problems principally arise from lack of lexical or syntactic knowledge. For instance, if a learner encounters a lexical item the meaning of which the learner does not know, he or she has to guess the meaning based on contextual information. Reading such a sentence should take more time than reading a sentence without unknown lexical items. Given this, some learners' problems can be identified by measuring his or her