Orthographic Processing of Proper Names: A Proposal to Investigate the Orthographic Cue for Second Language Readers

Kimberly Klassen
Tokyo Woman's Christian University

Abstract

A standard treatment of proper names in second language (L2) vocabulary analyses is to categorize them as known items. This treatment is often supported by the assumption that the form of the proper name (i.e., the initial capital letter) and the context will indicate to the L2 reader that the item is a proper name. The aim of this work-in-progress report is to propose a methodological approach for investigating the claim that form facilitates proper name recognition for L2 readers, in particular for those whose first language (L1) employs a different orthography and writing system than English. While previous research has looked at the orthographic processing of proper names for L1 users, a gap exists in the literature for L2 readers. The assumption that L2 readers are skilled at identifying proper names from the initial capital warrants investigation as there are implications for the processing load in both pedagogical and testing situations. A proposed method to investigate the initial capital letter as a cue for L2 English proper name recognition is a name decision task, a variation on a lexical decision task, using English proper names and common nouns that are familiar to Japanese participants. Reaction times and error percentages to proper names with case type as a factor (i.e., initial capital letter or all lowercase) will be submitted to a linear mixed effects analysis.

Keywords: proper names, second language reading, orthographic processing, orthographic cue, lexical decision task

1 Introduction

This study presents a work in progress that aims to investigate whether the form of English proper names (i.e., the initial capital letter) facilitates name recognition in second language (L2) readers of English whose first language (L1) has a different orthography and writing system than English. It has been assumed that proper names are easily identified and understood by L2 readers: the form and context will signal that the item is a proper name (Hirsh & Nation, 1992). This assumption has been cited in subsequent research to support the treatment of proper names in lexical analyses as known or low-burden items (e.g., Hu & Nation, 2000; Nation, 2006; Webb & Rodgers, 2009). Because proper

names can make up a significant portion of the items in a text, for example, 4–5% of fiction texts (Francis & Kucera, 1982), assuming proper names are known items can have a significant impact on the results of text coverage counts.

There is, however, little empirical support for the assumption. A study that did investigate the contextual aspect of the assumption is Klassen (2021), which looked at how well L2 readers were able to use sentential context to identify proper names and found that they were successful in only about a quarter of cases. This finding suggests that context might not always be helpful for L2 readers to identify proper names as such. As for the second part of the claim, that the form of proper names facilitates recognition for L2 readers, there is no study (that the author is aware of) that has investigated this claim. The work in progress described here aims to fill that gap.

2 Orthographic Processing of Proper Names

An area that has not received much attention in L2 reading research is orthographic processing (Nassaji, 2014). Since L2 readers arguably have less print exposure than L1 readers, it seems informative to look at the role of orthographic processing in L2 proper name recognition. There has been research into the orthographic processing of L1 proper names. A seminal study, Peressotti et al. (2003), used the lexical decision task paradigm to investigate recognition of L1 Italian proper names and common nouns. Like English, Italian proper names are capitalized, while common nouns are not.

Prior to Peressotti et al. (2003), studies that investigated the processing of these two word classes suggested that processing differences occur at the semantic level. No previous studies had suggested that differences in processing might occur at the orthographic level, even though there are orthographic differences between the two classes in many languages (i.e., proper names are marked with an initial capital letter). Therefore, their aim was to investigate if there is an effect of orthography on the visual recognition of proper names and common nouns for L1 Italian users. Peressotti et al. (2003) pointed to theories of abstract letter identities (e.g., Coltheart, 1981) which propose that word recognition is not affected by size, type, or case of letters, and thus would predict that processing of names and common nouns is not affected by the case of the initial letter. In contrast, episodic theories (e.g., Logan, 1988) suggest that word recognition is facilitated by earlier episodes with target words, and therefore would predict faster response times to proper names with initial capital letters and common nouns with lowercase letters.

Peressotti et al. (2003) tested these predictions by conducting five experiments. The sample (N = 173 over five experiments) was composed of L1 Italian adult participants. In experiments 1, 2, and 3, participants were shown sets of proper names, common nouns, and pseudo-words (derived from the real words by altering one or two letters). In the first two experiments, the stimuli were presented in one of two case conditions: first letter capitalized (e.g., Giovanni, Factory) or all letters capitalized (e.g., GIOVANNI, FACTORY). In experiment 3, stimuli were presented in either all lowercase letters (e.g., john, friendship) or only initial letter capitalized (e.g., John, Friendship). Experiment 4 was an auditory experiment.

In experiment 5, illegal non-words (random strings of consonants) were used instead of pseudo-words. The proper names and common nouns were matched for frequency and length. Regarding proper names for people, the authors focused their inquiry on generic names, those that can refer to an actual or potential person, rather than those names that refer to specific, famous people. The task was a lexical decision (i.e., to indicate whether the target item is a word or not). Participants were instructed to respond to the lexical decision task irrespective of the case of the letters.

The authors found that proper names with an initial capital letter are recognized faster than common nouns with an initial capital letter, and faster than proper names and common nouns with lowercase first letter. Proper names and common nouns with the first letter in lowercase are recognized at the same speed as common nouns with an initial capital letter. The findings indicate a "facilitation effect for proper names with the first letter capitalized" (p. 106). There was no advantage for proper names in the auditory condition, which suggests the effect occurs at the orthographic level, not semantic. Moreover, no effect was found for proper names in the lexical decision task with illegal non-words, which suggests that the initial capital letter is important only for lexical processing.

Peressotti et al. (2003) suggest that no current theoretical account can fully explain these findings. That the form of the initial letter affects lexical access by speeding up proper name recognition seems to provide counterevidence for the orthographic representation of abstract letter identities theory (Coltheart, 1981), which predicts that identification of a letter string is independent of case. However, their data do not support the episodic theory (Logan, 1988) either. Common nouns with an initial capital letter and proper names with an initial lowercase were *not* recognized more slowly than common nouns with the first letter in lowercase. Therefore, the familiarity of the stimuli (the case form that we normally see) did not affect response times. In their conclusion, the authors propose that their findings can be explained under abstract letter identities theory: "the upper-case lower-case distinction is abstract in nature as it is an intrinsic property of letters" (Peressotti et al., 2003, p. 108). Thus, they propose that the initial capital letter is a crucial orthographic cue that can speed up word recognition. Also, they claim that letter case contains relevant orthographic information because it can distinguish between word classes.

Peressotti et al.'s (2003) investigation has important implications with respect to the processing of L2 proper names. As summarized above, their findings suggest that the initial capital letter facilitates proper name recognition for L1 users. It is interesting to consider whether this facilitation effect also exists for L2 users. It is not known whether the initial capital letter facilitates proper name recognition for L2 users whose L1 uses a different writing script or orthography. For example, in German, all nouns are capitalized, while in English, only proper names and related adjectives are capitalized.

Proficiency level may also be a variable for how efficiently L2 readers orthographically process proper names in continuous text. There is evidence that as L2 proficiency increases, orthographic knowledge is used more than phonological knowledge (e.g., Chikamatsu, 2006; Nassaji, 2003). Thus, it might be the

case that L2 users with limited processing experience do not process uppercase and lowercase letters as efficiently and automatically as L1 users. For example, letters that look similar in both cases (e.g., Cc, Oo, Zz) might be more difficult to distinguish in continuous text than letters that are dissimilar (e.g., Aa, Bb, Ee). As Alderson (2000) notes, "Since difficulty in processing letters is related to automaticity of word identification, and since speed of word recognition affects speed and efficiency of reading, one might expect that second language readers processing different orthographies or scripts might experience greater difficulty" (p. 75). Thus, the assumption that L2 readers can easily identify proper names in continuous text by the initial capital letter needs to be confirmed with empirical data. This might be particularly relevant for readers whose L1 has a different writing script than English.

To investigate whether the initial capital letter facilitates recognition of English proper names for L2 readers with a different L1 writing system than English, an experiment is proposed: The participants' task will be a name decision task (i.e., *Is the stimulus a name? yes or no),* a variation of the lexical decision task. The research question is: To what extent does the initial capital letter on English proper names facilitate name recognition for L1 Japanese L2 English participants? While the initial capital letter facilitates recognition of proper names for L1 users, it is hypothesized that this facilitation effect will not be seen in L1 Japanese L2 English readers due to the fact that they have less processing experience than L1 English readers. Moreover, given the differences in how proper names are marked in the two languages (i.e., English with an alphabetic writing system and Japanese with a morphosyllabic and syllabic writing system), it is predicted that L1 Japanese L2 English participants will be less sensitive to the initial capital letter on English proper names.

3 Proposed Methodology and Data Analysis

3.1 Participants

Participants will be 20 Japanese university students at an intermediate proficiency level (B1 on the Common European Framework of Reference for Languages [CEFR] scale). As noted above, orthographic knowledge may be used more than phonology as L2 proficiency increases, so it is important that proficiency level is controlled for.

3.2 Materials

A set of 80 English proper names will be selected, and a list of corresponding common nouns will serve as distractors in the name decision task. (A list of pseudo-words will not be generated for the purpose of the experiment). The proper names and common nouns will be matched for frequency and length. The frequency of proper names will be determined from a corpus of Japanese junior and senior high school English textbooks. Common nouns will be selected from high-frequency word lists (i.e., first 3,000 words of the Corpus

of Contemporary American English [COCA] corpus). Before the list of items is finalized, in order to confirm that the proper names and common nouns are familiar to participants, six Japanese informants from the same population as the participants will be asked to rate the items on a familiarity scale. To control for the effects of semantic category, proper names will be selected from categories of people and places, and common nouns will also be limited to two categories, as was done by Peressotti et al. (2003) in their study. Items that have both a common noun and a proper name usage (e.g., *Rose, rose*) will not be used.

Target items will be presented in one of two conditions: with an initial capital letter or all lowercase. Two lists will be created in a Latin Square design, whereby if an item appears with an initial capital letter in one set, the same item will appear in all lowercase in the second set (e.g., *Michael, teacher; michael, Teacher)*.

The set of 80 target proper names used with 20 participants will achieve 1,600 word observations per condition (i.e., initial uppercase or all lowercase). Because reaction time experiments investigate small effects, 1,600 observations per condition are needed to ensure that the experiment is properly powered (Brysbaert & Stevens, 2018).

3.3 Procedure

The participants' task will be to decide whether a target item is a name or not (i.e., *Is this a name? yes/no*). Participants will be told that there are two categories of proper names: people's names and place names. They will be instructed to answer irrespective of the letter case. Accuracy and speed will be emphasized in the instructions. Participants will be trained with 20 items (i.e., 10 proper names and 10 common nouns), which are different than the items used in the experiment.

3.4 Data Analysis

Only responses to the proper names will be analyzed. Incorrect responses and reaction times beyond the limit value (two standard deviations plus or minus the mean) will be excluded from the analysis. Mean correct reaction times and error percentages will be reported for one fixed factor: case type (initial capital letter and all lowercase). These data will be submitted to a linear mixed effects analysis with R (R Core Team, 2022) with case type as a fixed effect, and participants and items as random effects. According to Brysbaert and Stevens (2018), mixed effects analyses are more powerful than traditional separate F1 (participants) and F2 (items) analyses. Because there are two random variables (i.e., participants and items), the authors contend that it is preferable to employ a single analysis that takes both into account.

4 Summary

While it has become standard in L2 lexical analyses to treat proper names as known vocabulary under the assumption that the form and context will provide relevant information to the L2 reader, little research has been done to investigate

this assumption. The proposal described in this study aims to investigate the claim that the initial capital letter serves as a cue for L2 English proper name recognition for readers whose L1 uses a different orthography than English. A name decision task experiment is proposed using a set of familiar proper names with case type (i.e., initial capital letter or all lowercase) as a factor; common nouns are used as distractors for the purpose of the task. Reaction times and error percentages for the proper names will be submitted to a linear mixed effects analysis.

References

Alderson, J. C. (2000). *Assessing reading.* Cambridge University Press.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 1–20. https://doi.org/10.5334/joc.10

Chikamatsu, N. (2006). Developmental word recognition: A study of L1 English readers of L2 Japanese. *The Modern Language Journal, 90*, 67–85. https://doi.org/10.1111/j.1540-4781.2006.00385.x

Coltheart, M. (1981). Disorders of reading: Their implications for models of normal reading. *Visible Language, 15*, 245–286.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage.* Houghton Mifflin Company.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696.

Hu, M. H., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Klassen, K. (2021). Second language readers' use of context to identify proper names. *The Reading Matrix, 21*(2), 85–105.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal, 87*, 261–276. https://doi.org/10.1111/1540-4781.00189

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching, 47*(01), 1–37. https://doi.org/10.1017/s0261444813000396

Nation, I. S. P. (2006). How large a vocabulary size is needed for reading and listening? *The Canadian Modern Language Review, 63*, 59–82. https://doi.org/10.3138/cmlr.63.1.59

Peressotti, F., Cubelli, R., & Job, R. (2003). On recognizing proper names: The orthographic cue hypothesis. *Cognitive Psychology, 47*, 87–116. https://doi.org/10.1016/S0010-0285(03)00004-5

R Core Team. (2022). R*: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*, 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x