The detriment that error production creates is affected by non-L1 speakers' linguistic

group membership

Abstract

We present three studies that investigate the effect of group-level language ability expectations on language ability judgements. Study 1 identifies expected English-language ability levels that native English speakers' have for a number of non-native English-speaker groups. Based on the results, two text-based written-guise studies were conducted investigating the level of detriment that grammatical and lexical/typographical errors created on English-language ability ratings for different author guises (Swedish, Chinese, English) in formal (Study 2) and informal contexts (Study 3). In both contexts, grammatical errors produced by the guise representing the lower-ability non-L1 group were overlooked, while the same errors produced by the other guises significantly lowered the ability ratings. Our results coincide with the idea of the 'sympathetic native speaker' and expand it, suggesting that expected language level based on linguistic group membership inversely affects the level of sympathy/tolerance demonstrated. We link this to possible pedagogical implications.

The detriment that error production creates is affected by non-L1 speakers' linguistic group membership

1. Introduction

English is a lingua franca spoken by 1.3 billion people around the world as a non-L1 (Eberhard et al., 2020).[1] Even though some of these speakers have a native-like proficiency in their two (or more) languages, many non-L1 English-speakers' are not considered native-like in English, particularly if they have learned English after childhood (e.g., Johnson & Newport, 1989; Schmid et al., 2014; Spillner, 2017). Thus, in urban multicultural contexts, native speakers of English are likely to have contact with non-L1 speakers representing a range of English language abilities from beginner to (near-)native.

Even though a pluralist approach to L2 language use is often adopted by linguists (including the authors), it is widely reported that speakers having a lower than native-like proficiency is not without social consequences. There has been significant research for decades on attitudes and perceptions towards individuals' language/dialect (see Dragojevic et al., 2021 for a review) and it is known that non-L1 speakers language use affects how they are perceived by others (Delamere, 1996; Dragojevic et al., 2017; Derwing et al., 2002; Mirshahidi, 2016; Tajfel & Turner, 2004) - in a similar way that native speakers make judgements on other native speakers based on their accent or other linguistic characteristics (e.g., Levon et al., 2021). Speakers with a non-L1 accent are for example judged more negatively by native speakers with regards to status and social rankings in comparison to native speakers (e.g. Garrett, 2010; Giles, 1970; Lambert et al., 1960), and when it comes to written production of language, perceived errors in non-L1 speakers' texts have a negative effect on judgments of the author's trustworthiness, friendliness and competence (Planken et al., 2019) potentially putting non-L1 speakers in a disadvantaged position in a number of

written contexts (e.g. job applications, online dating websites, social media). Our research looks at perceptions from a slightly different perspective and investigages if expectations of language ability at the group level impact language ability judgements.

L1 speakers have been found to be sympathetic towards non-L1 speakers' linguistic errors when making assessments on their linguistic abilities (e.g., Galloway, 1980). Namely, when assessing non-L1 speakers' target language knowledge, L1 speakers are more lenient towards non-L1 than L1 speaker's errors, in particular if the non-L1 speaker produces many errors, thus demonstrating low target language ability. That is, native speakers can be lenient towards errors based on the language ability an individual demonstrates. However, as far as we are aware, the questions as to whether native speakers' language ability expectations associated with different linguistic groups affect language ability ratings has not been widely studied, in particular outside the class-room context. The current paper aims to fill this gap and seeks to answer the question 'Does ethnolinguistic group membership of the writer affect native judge acceptability ratings of writing errors?'. We do this by investigating grammatical and lexical/typographical errors in formal and informal context on the ratings of English language ability of non-L1 speaker guises representing two linguistic groups – Swedish and Chinese – and a native speaker guise, aiming to shed light onto potential implicit ethnolinguistic bias.

1.1. Errors

Even though it can be difficult to define what an error is due to factors including ongoing language change (Appelman & Schmierbach, 2018), dialectal differences (Bender, 2005; Wheeler et al., 2012), different communicative modalities (spoken, written, or even signed; Derwing et al., 2002; Queen & Boland, 2015) and registers (e.g., Cargile, 1997) native speakers, including children, have a high level of awareness as to what is an acceptable/good

sentence in their native language and can give accurate judgements on sentences (e.g., Ambridge, et al. 2020; Bialystok, 1986), indicating that norms for goodness/acceptability do exist and speakers are aware of them. In the current paper, we will focus on errors in English texts which are not associated with a particular sociolinguistic/regional dialect but are relatively clear instances of typos, omissions or incorrect use of grammatical functions, or muddled complex sentences (e.g., where the latter part of the sentence does not grammatically follow from the earlier part) and investigate their effect on judgements of the writer's English language ability, while manipulating the writer's assumed linguistic group membership.

1.1.1. Native speakers

It is well documented that native speakers' language, either written or spoken, is not error free and that they produce grammatical errors (Kantz & Yates, 1994; Queen & Boland, 2015) and lexical/typographical errors (Wilcox et al., 2014; Figueredo & Varnhagen, 2005) This is partly due to (a) native speakers not being a homogenous group, differing, for instance, in vocabulary size (e.g., Nation, 2006) and grammatical knowledge (e.g., producing *What's* Xs or *There's* Xs when referring to plural complements, Mumford, 2009, p. 139; the use of the passive voice, Dąbrowska & Street, 2006) and (b) processing, planning and attentional issues related for example to a high level of cognitive load, tiredness or linguistic interference resulting in 'slips' (e.g., Foster, 2007) (e.g., accidentally writing *their* for *there*).

1.1.2. Non-L1 speakers

Non-L1 speakers also produce a range of errors. Similarly to native speakers, non-L1 speakers can produce errors due to processing, planning and attentional issues (e.g., accidentally producing *their* for *there*). In addition, some of the errors reflect the speaker's

level in the target language (i.e., using/spelling words incorrectly or making errors in grammatical constructions such as the passive). A non-L1 speaker's native language characteristics are also likely to have an effect on the errors produced due to transfer and linguistic distance between language groups (e.g., Spillner, 2017) (e.g., a native Finnish speaker might erroneously omit articles or a native Japanese speaker might erroneously omit plural -s in L2 English due to Finnish not having articles and Japanese not having a universal plural suffix). The speaker's linguistic, educational and cultural background can likewise have an effect on the group's overall level of the target language (e.g., in relation to spelling, vocabulary size, knowledge of grammar).

1.1.3. Errors in written texts

English-speaking adults can (and often do) base their assessment of written texts on Standard English norms, often implicitly judging the writer's lexical, typographical and grammatical choices. However, not all errors are equally salient. A number of studies report that lexical/typographical errors appear to be more detrimental to the text quality judgements than grammar-related errors (Beason, 2001; Hairston, 1984; Kantz & Yates, 1994; Leonard & Gilsdorf, 1991; Queen & Boland, 2015; Roberts and Cimasko, 2008; Santos,1988). Furthermore, different error types seem to affect the perceived characteristics of the writer differently (Derwing et al., 2002; Kreiner, Schnakenberg, Green, Costello, & McClin, 2002; Queen & Boland, 2015). Queen and Boland (2015) distinguished between 'typos' (a surface mechanical error) and 'grammos' (an error of grammar) and their results suggested that both types of errors influenced perception of the writer's native-speaker academic qualities, but only 'grammos' affected readers' social perceptions of the writer. Kreiner et al. (2002) conducted three experiments investigating native-speakers' rating of spelling errors and the effect those errors had on the judges' perception of the writer. They found that a large

number of errors in the short essay resulted in the judges forming a more negative perception of the writer and assuming lower writing ability. Some obvious spelling errors also significantly affected the ratings of intellectual ability.

Previous work has also demonstrated that judges have different expectations between formal and informal written contexts (Cargile, 1997; Mozafari, El-Alayli, Kunemund, & Fry, 2019). Mozafari et al.'s (2019) study sought to investigate the influence of spelling and grammatical mistakes on the inclination to use the services of blue-/white-collar businesses, effectively looking at how socioeconomic- and education-related expectations influenced error perception. Results indicated that potential buyers were less inclined to use white-collar businesses when their adverts had errors, while blue-collar businesses were not as affected. They suggest this was because the judges had different expectations of education level between blue- and white-collar employees. This indicates that the reader's expectation of a speaker's language knowledge at sociolinguistic group level in written texts affects the level of detriment that spelling/grammar mistakes create.

In summary, previous research has shown that errors in written texts can affect listeners'/readers' judgements on the writer. Within these instances, the error type, the register of the context (i.e., formal vs. informal) and the expectation of ability based on educational background/socioeconomic status can have an effect on the perceived level of detriment that errors create. The current study expands this literature and investigates if language ability judgements are affected by the writer's linguistic group membership.

1.2. Personal names and group membership

Native speakers are likely to associate personal names that do not correspond to the typical names in a given society with minority ethonolinguistic groups (Birkelund et al. 2014; Greenwald, McGhee & Shwartz, 1998), and can make judgements for/against individuals

based on personal names, for example in the housing market (e.g., Carpusor & Loges, 2006) and in job applications (e.g., Cotton, O'Neill & Griffin, 2008).

The current study uses pseudonyms of individuals with names indicating minority ethnolinguistic group membership in the context of the UK, which is likely to be a relatively strong indication that the author is not a monolingual native-speaker of English. However, in contrast to the often negative biases found above, previous research has identified that second language users might, in fact, benefit from their distinct names as native speakers often display an increased tolerance of inaccuracies made by second language learners, otherwise termed the *sympathetic native speaker* effect.

1.3. Sympathetic Native Speaker

Even though it is often the case that errors increase negative attitudes towards a speaker/writer, be it in relation to their personality, competence, intelligence, education level, friendliness, group membership, or other traits (e.g., Dragojevic et al., 2021; Garrett, 2010), it has been suggested that in language ability judgements, native speakers might be more 'sympathetic' towards low ability language learners' than high ability learners' errors (Davies, 1983; Fayer & Krasinksi, 1987; Galloway, 1980; Preston, 1981) . Galloway (1980) investigated the attitudes of native and non-native Spanish speakers towards non-L1 Spanish learners ranging from intermediate to near-native. She found that while the native speakers were more tolerant of errors than their non-native counterparts, both groups were more critical of those non-L1 speakers who spoke both fluently and accurately. Studies into L2 speech acquisition and assessment often refer to a 'sympathetic native speaker' in their frameworks on oral interactions (Frank, 2010; Ayhan & Uğur Türkyılmaz, 2015) where lower competencies of language production are expected to be understood by a sympathetic native speaker (a person willing to spend more time and effort with the L2 learner) but not

understood by a general audience. A further study considered L2 identity development within an online context and mentioned the 'sympathetic native speaker' (Klimanova & Dembovskaya, 2013). However, the sympathetic native speaker effect was not the focus of these studies.

If the sympathetic native speaker effect is related to the perception of the speaker's language ability, it could then entail that native speakers also view other native speakers' errors more critically than the same errors produced by non-L1 speakers because a native judge (assuming a prescriptivist outlook) may expect relatively error-free language from a native author/speaker in comparison to at least some non-L1 speakers. Consequently, different standards may be applied to the native and non-native speakers' language production whereby certain errors might be overlooked in interlanguage so long as successful communication is achieved (see Davies, 1983; Derwing et al., 2002; Schmitt, 1993), while accuracy is the expectation of a native speaker. Rubins and Williams-James' (1997) study investigating the impact of ethnolinguistic stereotypes on composition evaluations by academic instructors on native-speaker-written essays that had native and non-native pseudonyms supports this suggestion. They found that the native speaker instructors were more lenient on essays with a Thai name and more critical in their feedback to 'students' with Danish or Anglo-American names. One explanation put forward by Rubins and Williams-James was that U.S. native English speakers were marked more harshly because "they should have known better" (1997, p.150).

The effect being ability-related would entail that a high-ability non-L1 speakers' texts would also be judged harshly, in the same/similar way to a native speaker (i.e. they should have known better) because a high level of language ability is expected, although the non-native status might result in more sympathy even with high level ability in comparison to native-speakers.

To summarize, previous research suggests that native speakers can be sympathetic towards non-L1 speakers' linguistic errors, in particular if the learner demonstrates a low ability level and that native-speakers can show different levels of sympathy towards different speaker groups. This phenomenon is understood outside of research contexts, as both the CEFL (Common European Framework of Reference for Languages) and ACTFL Guidelines (American Council on the Teaching of Foreign Languages) in their older renditions acknowledged how a sympathetic native speaker facilitates lower level L2 language learners and is used in their respective assessment frameworks. Current guidelines as for 2023 for both foundations now identify 'sympathetic interlocutor' (CEFL, 2020) or a 'sympathetic audience (ACTFL, 2015). However, as far as we are aware, no previous study has investigated if the sympathetic native speaker effect can be affected by the language level expectation that the native-speaker judge has about the non-L1 speaker at ethnolinguistic group level, tested outside the classroom when the linguistic group membership of the supposed writer is manipulated. That is, we investigate if native speakers demonstrate a higher level of sympathy if they expect the non-L1 speaker to have a lower language level based on the writer's linguistic background. The current study will investigate this question by using the written guise technique and general public (i.e., not educator) native speaker judges.

2. The present study

Native and non-native speakers produce grammatical and lexical/typographical errors/slips in spoken and written language and error production can have an effect on listeners'/readers' perception of the speaker/writer or of their abilities. It is also possible that a speaker's/writer's expected language abilities affect native speakers' judgements.

Investigating formal (Study 2) and informal written contexts (Study 3), the present study explores whether grammatical and lexical/typographical errors create similar detriment in language ability judgements, whether the formality level affects ratings, and whether the speaker's linguistic group membership affects how sympathetic the native speaker judge is towards the speaker's grammatical and lexical/typographical errors.

To inform Study 2 and Study 3, a preliminary study (Study 1) investigates if non-L1 speakers' self-reported collective English-language ability levels (Office for National Statistics, 2016) correspond to British native-English speakers' perception of the ability level of different non-L1 groups. We were particularly interested in the native speakers' English-language expectation for two linguistic groups, speakers of one of these groups representing a systematically, high English-language ability (Swedish) and one representing a more mixed English-language ability (Chinese).[2] These two languages were chosen for the following reasons. First, the Census data reports that while 98-99% of Swedish, Danish, Finnish and German native/main language speakers state that they can speak English 'well' or 'very well' (the two highest categories available), only 38-76% of Gypsy/Traveller languages,[3] Pakistani Pahari (with Mirpuri and Pothwari), Vietnamese, Yiddish, Cantonese Chinese, Mandarin Chinese and all other Chinese speakers selected either of those two options. In other words, approximately 25-60% of the latter native speakers stated that they could speak English 'not well' or 'not at all.' Thus, when native English-speakers interact with native/main language Swedish speakers, they are likely to observe nothing but a high level of English use, but when interacting with native/main language speakers of any variant of Chinese, their experience of English language ability is likely to be more mixed. This is likely to result in a general assumption that native Swedish speakers as a group have higher English language ability than Chinese speakers as a group. Second, the number of Swedish (19k) and Cantonese (43k), Mandarin (21k) and all other Chinese speakers (135k) in the UK is

substantial compared for example to Finnish speakers (approximately 7k) (Office for National Statistics, 2016), suggesting that British people are more likely to have encountered Swedish and Chinese speakers due to the population size.[4] Third, due to the presence of Swedish and Chinese names in the UK as a result of popular culture, politics, history, sport and retail outlets we assumed that typical names from these linguistic backgrounds (compared, e.g., to Finnish, Gypsy/traveller or Yiddish names) would be relatively familiar to British speakers, allowing accurate perception to be associated with the guises in Study 2 and Study 3.

Our predictions are the following: (i) based on the idea of the sympathetic native speaker, if the speaker's expected language level affects the English ability ratings, errors will differentially affect the perceived language ability such that errors should be rated more detrimental in the high language ability level guise (Swedish) and in the native speaker guise (English) than in the low language ability level guise (Chinese); (ii) lexical/typographical errors are likely to be more salient to readers than grammatical errors and thus will increase the detriment level (e.g., Brandenburg, 2015; Queen & Boland, 2015); and (iii) the formal context should increase the level of detriment that errors create.

2.1. On the selection of judges

The current study uses native English-speaking judges to see if linguistic group membership (ENG, SWE, CHI) affects the level of detriment in perception of error production in written texts created on the author's language ability. It is important to highlight that we do not assume that native speakers of English are the only group whose perception of language ability is important/informative. Given that English is a Global Lingua Franca, native and non-native speakers' reactions to errors in English texts are important in real life writing (and spoken) interaction. However, in the current study, we chose native English speaker judges as

we wanted to control for the judges' linguistic background to avoid noise in the data due to different native speakers' potentially different expectations of different nationalities' English language abilities.

3. Study 1 – native speaker's expectation of non-L1 speaker's English language ability

The Office for National Statistics' (2016) Census data show that, in the UK, almost all speakers whose first/stronger language is Swedish assess their English-language level as 'good' or 'very good', while only 61%-76% of Chinese speakers give this self-rating. This can affect the level of English-language expectation that British native-English speakers have for these two linguistic groups. To ascertain that these self-ratings correspond to native British English speakers' perception of language ability at group level (i.e., that native speakers are aware of these differences in ability), we ran a survey asking native English speakers to rate the expected language level of different non-native speaker groups.

3.1. Method

We ran an online survey (via Qualtrics) with adult native English-speaking UK residents to investigate the perceptions of English language ability of speakers from nine commonly spoken minority language groups in the UK (Office for National Statistics, 2016). The nine languages selected represented a range of self-reported ability levels and different geographical regions of origin: Chinese (Mandarin/Cantonese), German, Polish, Punjabi, Spanish, Somali, Swedish, Tagalog/Filipino, and Welsh. The participants were asked to complete two tasks. First, they were asked to rank the nine speaker groups based on the speakers' expected English language ability (where '1' represented the highest ability group

and '9' the lowest). Second, participants were asked to provide a rating for each of the nine languages on the speakers' expected proficiency in English on a five-point scale: 'very low,' 'low,' 'intermediate,' 'high' and 'very high.' In both tasks, the languages were presented in randomized order for each participant to reduce response order bias.

3.2. Results and discussion

A total of 79 respondents completed the survey. For the first task of ranking the nine languages based on expected English language ability, Swedish ranked higher ($M = 3.00$; $SD = 1.50$) than Chinese ($M = 6.96$; $SD = 1.74$). The distributions of the ranking responses can be observed in Figure 1. In only five responses did Chinese rank higher than Swedish (between one and five ranks' difference), and of the 74 responses that ranked Swedish higher than Chinese, the average rank difference was 4.45 ($SD = 1.84$). A Wilcoxon Signed-Rank test showed that the overall difference between the ranking for Swedish and Chinese was highly significant ($z = 7.09$, $p < 0.001$).
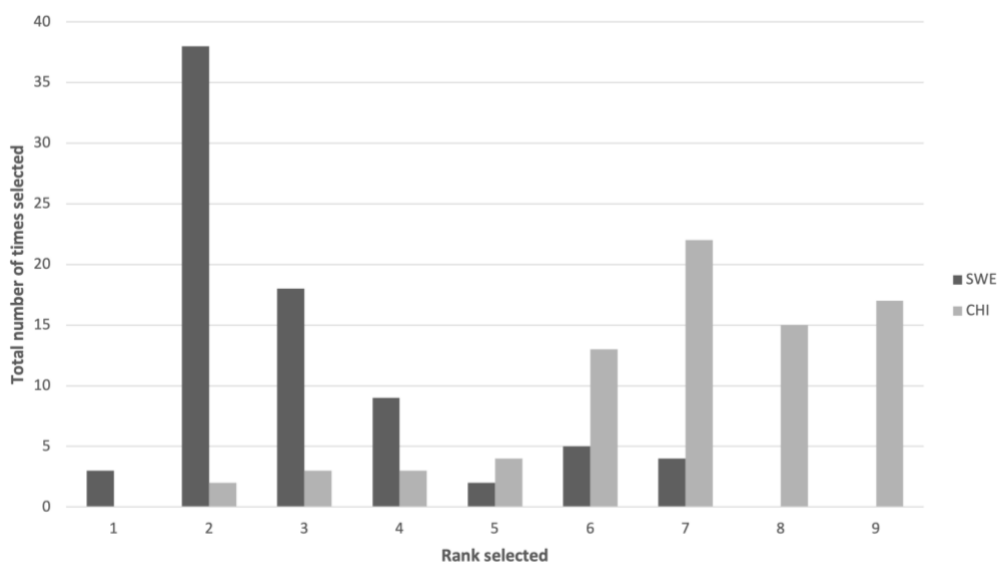
Figure 1. Rank distributions for Swedish and Chinese (Study 1).

For the second task, selections on the five-point scale from 'very low' to 'very high' representing the expected language ability of each of the nine speaker groups were converted into integer values (where 'very low' = 1 and 'very high' = 5). These scores were collected for Swedish and Chinese and the averages calculated. As in the first task, respondents ranked Swedish higher ($M = 4.22$, $SD = 0.78$) than Chinese ($M = 2.82$, $SD = 0.90$), with a significant difference observed between the two ($t_{(156)} = 10.38$; $p < 0.001$).

To recap, we found that the expected L2 English language level of Swedish speakers was higher than that of Chinese speakers, which corresponds to the Census language self-assessment English language ability data (Office for National Statistics, 2016). This offered a stable basis from which the next stage of the study could progress.

4. Study 2 – Judgments of errors in a formal text

4.1. Method

4.1.1. Respondents (Judges)

Two-hundred-and-eighty-nine (178 identifying as female; 111 identifying as male) native English speaker adults (henceforth 'judges') were recruited from members of the public in Bristol and Brighton, UK and people within a university in Bristol including staff, students and visitors. The participants ranged between 18-85+ years of age (age ranges were: 18-24 (n=85); 25-34 (n=54); 35-44 (n=68); 45-54 (n=52); 55-64 (n=15); 65-74 (n=13); 75-84 (n=1); 85+ (n=1)). Forty-seven disclosed that in addition to being a native English speaker they were also fluent in another language. None disclosed of having known severe cognitive deficits. None had taken part in Study 1.

4.1.2. Texts used

An error-free job cover letter was created, from which two further letters were generated, one with eight grammatical errors and the other with eight lexical/typographical errors.[5] The errors were relatively subtle in that they did not impede meaning and thus could have been produced by an L1 speaker or a high-level non-L1 speaker.

Adopting some of the error types reported in previous literature (e.g., Bender, 2004; Wilcox et al.'s 2014), the grammatical errors consisted of four agreement errors concerning plurality and tense; the use of a resumptive pronoun; an article omission; a copula omission; and a redundant use of a derivational suffix (i.e. were similar to errors reported in English speaking classrooms). The lexical/typographical errors consisted of typical and slightly less typical errors in the form of eight words used (*their* for *there*, *proud* for *pride*, *irregardless* for *regardless*, *passed* for *past*, *affective* for *effective*, *except* for *accept*, *greatful* for *grateful* and *that* for *than)*.

The two error texts and the control text were as closely matched for length and structure: both started with 'Dear Mr. Smith' and ended after 'Yours sincerely;' the lexical/typographical error text and the control text contained 251 words, and the grammatical error text contained 247 words (due to some of the grammatical errors being omission of words).[6]

4.1.3. The Guises, i.e., the linguistic background of the writer of the text

Based on self-assessment in the 2011 UK Census (Office for National Statistics, 2016) and our Study 1, we selected a high English language proficiency speaker (Swedish, henceforth SWE) and more mixed English language proficiency speaker (Chinese, henceforth CHI) and a native English (henceforth, ENG) speaker as our guises. We chose appropriate names to correspond the guises so that the first and the last names were amongst the 10 most common

female first names in the guise's native country and showed no overlap in use between two countries (e.g. the name *Alice* is common in England and Sweden and was thereby excluded). The most popular name was not always selected to avoid a stereotypical/satirical name.[7] The names chosen were: Sarah Wilson (ENG), Astrid Karlsson (SWE) and Li Xiu Yin (CHI).

We created a copy for each text (grammatical-error; lexical/typographical-error; and control, that had no errors) for each guise. That is, every cover letter was signed by one of the three guises, the name indicating the writer's linguistic background. For every writer condition (ENG, SWE, CHI) the texts remained exactly the same except for the change of name, thus allowing us to establish if the different guise resulted in different judgements on the quality of the text. That is, we had a 3 (text error type: grammatical, lexical/typographical, control) x 3 (linguistic background: ENG, SWE, CHI) design. Judges were randomly allocated to one of the nine conditions.

4.1.4. Questionnaire

We created a questionnaire consisting of five Likert scale questions ranging between 1 (poor English) and 9 (perfect English). Judges were asked to give a score on the text by circling a number, to give answers to each of the five questions, and to circle only one number per question. Two of these five questions asked about the judge's perception of the writer's overall proficiency (Command of English; Naturalness) and three about the writer's lexicogrammatical abilities (Grammar; Vocabulary; Spelling). The Likert scales were followed by an optional question that allowed the judges to leave qualitative commentary on their opinion of the text and why they scored it the way they did.

As this study investigated perceptions of language proficiency, to avoid the judges noticing that we were getting ratings for different linguistic groups (ENG, SWE, CHI), we

used a between-participants method (i.e., the judges only ever read one text and completed one questionnaire).

4.1.5. Procedure

The data were collected face-to-face using paper questionnaires. Researchers and research assistants set up a small stall in various locations (e.g., local park, university campus, school fair) and asked passers-by to complete a questionnaire. If the passers-by agreed to take part, the researchers then explained that the study explores what English speakers think is 'good English' by asking the participant to read a short text, give as accurate ratings on the text as possible, and answer some demographic questions. Reading the text and completing the questionnaire took approximately 3-5 minutes.

4.1.6. Coding

The responses from the questionnaires were transferred into an Excel spreadsheet. If a judge had circled more than one number on the Likert scale (<10 instances) without crossing out the redundant responses, a decision was made to always take the highest number circled as the response.

4.2. Results

The results of the five dependent variables stemming from the questions asked in the survey were combined (i.e., mean values calculated) to each participant to create the composite variable of 'Overall Judgement' (Cronbach's α of internal consistency = 0.88).

A factorial ANOVA was conducted to compare the effect of 'Overall Judgement' on Text type (Grammatical error, lexical error, control) and Nationality (ENG, SWE, CHI) with a significant effect observed for 'Text Type' (F (2, 280) = 31.32, $p < 0.001$, $R^2 = 0.20$). This

effect was investigated in greater detail using a Simple Effects Analysis comparing different 'Text types' per 'Nationality.' The estimated marginal mean values were calculated in each case, shown in Table 1 and Figure 2, followed by the results of the Simple Effects Analysis, shown in Table 2.

Table 1. Estimated Marginal Means for nationality and text type for Study 2

Text type	Mean	Std. Dev.	Std. Error	CI (95%)
English Control	6.83	1.32	0.24	6.36 to 7.30
English Grammatical error	5.72	1.28	0.23	5.27 to 6.17
English Lexical error	5.03	1.28	0.23	4.58 to 5.48
Swedish Control	6.62	1.32	0.24	6.15 to 7.0
Swedish Grammatical error	5.53	1.26	0.22	5.09 to 5.95
Swedish Lexical error	4.96	1.30	0.23	4.51 to 5.41
Chinese Control	6.34	1.32	0.24	5.87 to 6.81
Chinese Grammatical error	5.94	1.31	0.21	5.47 to 6.41
Chinese Lexical error	5.34	1.26	0.22	4.90 to 5.78
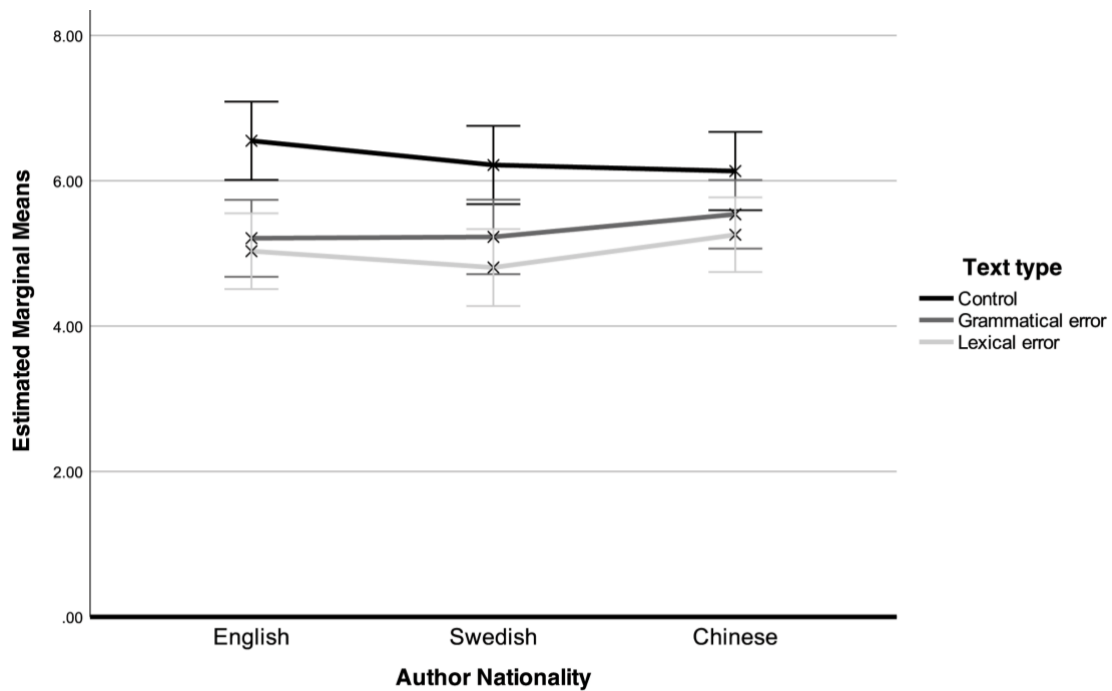
Figure 2. Plotted estimated marginal means for Study 2 (error bars = 95% CI).

As expected, all three Control (error free) texts were rated higher than their respective error texts, with no statistically significant difference across guises ($F_{(2)} = 1.45$, $p = 0.24$), suggesting that the participants noticed (at least some of) the errors and that when errors were not present the guise's linguistic background did not affect the judgements. The grammatical error text for the Chinese guise scored closer to its associated Control text ($M_{difference} = 0.40$), than the Grammatical error text for the Swedish ($M_{difference} = 1.09$) and English ($M_{difference} = 1.12$) guises. The difference between grammatical error and control texts was significant for the Swedish ($t_{(61)} = 3.59$; $p < .001$) and English guises ($t_{(59)} = 3.19$; $p = .001$), but non-significant for the Chinese guise ($p > .05$), suggesting that grammatical errors were largely overlooked in the Chinese author's text. The lexical/typographical error texts were judged similarly for the three guises (ENG $M_{difference} = 1.81$; SWE: $M_{difference} = 1.67$ CHI: $M_{difference} = 1.00$), with significant decrease in the judgements for the error texts in comparison to the control texts (ENG: $F_{(2)} = 14.16$, $p < 0.001$; Tukey HSD of $p < 0.01$ for Control vs. Lexical and Control vs. Grammatical Error; SWE: $F_{(2)} = 11.83$, $p < 0.001$; Tukey HSD of $p < 0.01$ for Control vs. Lexical and Control vs. Grammatical Error; CHI: $F_{(2)} = 5.96$, $p < 0.01$; Tukey HSD of $p < 0.01$ for Control vs. Lexical Error).

Table 2. Simple Effects Analysis for text types per nationality for Study 2.

Text type comparison	Mean Difference	Standardised effect size (d)	p
English Control vs. Lexical error	**1.81**	1.39	**<0.001*****
English Control vs. Grammatical error	**1.12**	0.85	**0.001****
English Lexical error vs. Grammatical error	**0.69**	0.54	**0.034***
Swedish Control vs. Lexical error	**1.67**	1.27	**0.001****
Swedish Control vs. Grammatical error	**1.09**	1.29	**<0.001*****
Swedish Lexical error vs. Grammatical error	0.58	0.45	0.074
Chinese Control vs. Lexical error	**1.00**	1.29	**0.002****
Chinese Control vs. Grammatical error	0.40	0.30	0.206
Chinese Lexical error vs. Grammatical error	**0.60**	**0.47**	**0.048***

* = weak; ** = strong; *** = very strong

4.3. Discussion

Our findings suggest that the Swedish guise's grammatical errors were judged similarly to the native English speaker guise - grammatical errors in these two guises resulted in significantly decrease in language ability ratings relative to the control texts, while the Chinese speaker's grammatical errors did not create a significant decrease in judgements. This means that the judges were more sympathetic towards the Chinese speaker than the Swedish and English speakers. Note that the control text judgements did not differ between the three guises, suggesting that the difference observed in the judgements for the error texts are not in any straightforward way related to differences in overall attitudes towards different linguistic/ethnic backgrounds. We argue that due to the expectation that the Swedish speakers' English language level is very high – similar to that of native English speakers, the errors produced by the Swedish and English guises were judged more harshly than those produced by the Chinese guise. However, speakers tend to be more lenient towards errors in informal contexts, such as housemate applications or job fliers (Mozafari et al., 2019; Queen & Boland, 2015) than in formal contexts, such as job cover letters. This registral and contextual distinction might influence the sympathetic native speaker effect that appears to be mediated by the non-L1 speaker's expected language level. To investigate this possibility, we conducted Study 3.

5. Study 3 - judgments of errors in an informal text

5.1. Method

The effect of errors in an informal written context (Facebook) was investigated to determine whether native speakers' judgements are different for the three guises in contexts where there is less expectation to produce error free or 'correct' texts.

5.1.1. Respondents (judges)

Three-hundred-and-eighty-four native English-speaking adults, henceforth 'judges' (239 identifying as female and 142 identifying as male; 2 as 'neutral' and 1 choosing not to disclose their gender), were recruited as per Study 2. The participants ranged between 18-85+ years of age (age ranges were: 18-24 (n=175); 25-34 (n=54); 35-44 (n=61); 45-54 (n=55); 55-64 (n=21); 65-74 (n=12); 75-84 (n=5); 85+ (n=1)). Sixty-eight disclosed that in addition to being a native English speaker they were fluent in another language. None disclosed of having known severe cognitive deficits, and none had taken part in Study 1 or 2.

5.1.2. Materials

To see if the same error types bring about different judgements when different levels of formality are associated with the context, the errors used in the informal context were matched as closely as possible to the ones used in the formal context (Study 2). Rather than creating one long Facebook post (corresponding in its length to the job cover letter in Study 2), to better mimic typical real life social-media posts, we created eight shorter Facebook posts for grammatical error, lexical/typographical error and control conditions. These were presented to the judges on 2 A4 sheets. The judges were instructed to base their judgements on the 8 posts collectively. Due to the difficulty in creating identical texts for grammar and lexical/typographical Facebook posts that would sound natural yet correspond to the errors in the formal text, two different texts were created, one for grammatical error and control texts

and one for lexical/typographical error and control texts. Participants were randomly assigned to read one of these 2-page Facebook texts.

To avoid formal or informal contexts to be judged differently due to significant differences in the length of the text or error distribution, the texts used were matched carefully. Both formal and informal texts had the same number of words (grammatical: 247; lexical/typographical 251 words). The word count for the informal text included hashtags but not emojis or the person's name. Both formal and informal texts had the same number of sentences (grammatical: 16; lexical/typographical: 16) and the same number of sentences had errors (grammatical: 6/16, lexical/typographical: 7/16).

The Facebook posts showed the profile name, which corresponded to the guise names used in Study 1 (ENG: Sarah Wilson, SWE: Astrid Karlsson, CHI: Li Xiu Yin). Mimicking real Facebook posts, each post was headed by the supposed writer's name and an avatar (a landscape), which was the same for all three guises.

5.1.3. Coding

The coding used the same protocol as in Study 2.

5.2. Results

As in Study 2, the five dependent variables were combined to create the composite variable of 'Overall Judgement' (Cronbach's α of internal consistency = 0.83).

A factorial ANOVA was conducted to compare the effect of 'Overall Judgment' on Text type (grammatical error and control; lexical/typographical error and control) and Nationality (ENG, SWE, CHI). A significant effect was observed for 'Text Type' (F (3, 370) = 34.92, $p < 0.001$, $R^2 = 0.24$). This effect was investigated in greater detail using a Simple Effects Analysis comparing different 'Text types' per 'Nationality.' The estimated marginal

mean values were calculated in each case, shown in Table 3 and Figure 2 followed by the

results of the Simple Effects Analysis, shown in Table 4. These comparisons did not include

combinations where the control text and the error text mismatched (e.g., Control (Lexical) vs.

Error (Grammatical)).

Table 3. Estimated Marginal Means for nationality and test type for Study 3.

Text type	Mean	Std. Dev.	Std. Error	CI (95%)
English Control (Lexical)	6.95	1.19	0.22	6.52 to 7.38
English Control (Grammatical)	6.72	1.15	0.20	6.33 to 7.11
English Lexical error	5.43	1.16	0.19	5.06 to 5.8
English Grammatical error	5.70	1.20	0.20	5.31 to 6.09
Swedish Control (Lexical)	7.40	1.16	0.22	6.97 to 7.83
Swedish Control (Grammatical)	6.70	1.15	0.21	6.29 to 7.11
Swedish Lexical error	5.63	1.14	0.19	5.26 to 6.00
Swedish Grammatical error	6.12	1.17	0.21	5.71 to 6.53
Chinese Control (Lexical)	7.32	1.19	0.22	6.89 to 7.75
Chinese Control (Grammatical)	6.73	1.15	0.21	6.32 to 7.14
Chinese Lexical error	5.79	1.15	0.20	5.40 to 6.18
Chinese Grammatical error	6.44	1.19	0.21	6.03 to 6.85
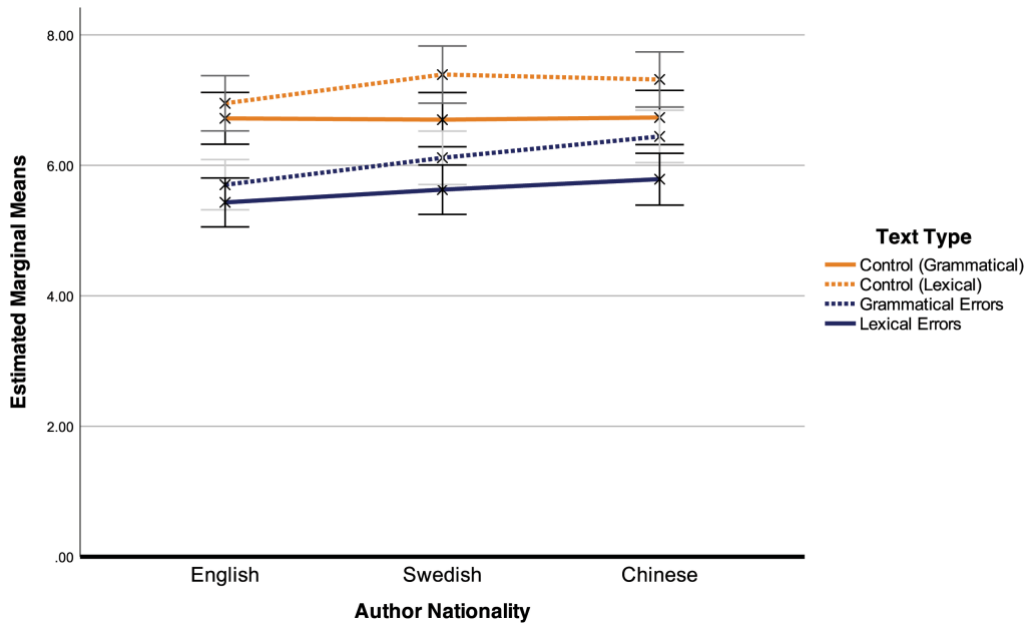
Figure 3. Plotted estimated marginal means for Study 3 (error bars = 95% CI).

Table 4. Simple Effects Analysis for text types per nationality for Study 3.

Text type comparison	Mean Difference	Standardised effect size (d)	*p*
ENG Control (Lexical) vs. Control (Grammatical)	0.23	0.20	0.435
ENG Control (Lexical) vs. Lexical error	**1.52**	**1.29**	**<0.001*****
ENG Control (Grammatical) vs. Grammatical error	**1.02**	**0.87**	**<0.001*****
ENG Lexical error vs. Grammatical error	0.27	0.23	0.323
SWE Control (Lexical) vs. Control (Grammatical)	0.69	0.60	0.025
SWE Control (Lexical) vs. Lexical error	**1.77**	**1.54**	**<0.001*****
SWE Control (Grammatical) vs. Grammatical error	**0.58**	**0.50**	**0.050***
SWE Lexical error vs. Grammatical error	0.49	0.42	0.086
CHI Control (Lexical) vs. Control (Grammatical)	0.58	0.50	0.054
CHI Control (Lexical) vs. Lexical error	**1.53**	**1.31**	**<0.001*****
CHI Control (Grammatical) vs. Grammatical error	0.29	0.25	0.326
CHI Lexical error vs. Grammatical error	**0.66**	**0.56**	**0.023***

* = weak; ** = strong; *** = very strong

All Control texts were rated higher than their respective error texts, suggesting the

participants had noticed (at least some of) the errors, and the differences between the Control

texts per nationality were not statistically significant. The text with grammatical errors for the

Chinese guise scored closer to its associated Control text ($M_{difference} = 0.29$) than the text with grammatical errors for the Swedish ($M_{difference} = 0.58$) and English ($M_{difference} = 1.02$) guises. The difference between grammatical error and control texts was significant for the Swedish ($t_{(59)} = 1.97$; $p = .05$) and English guises ($t_{(66)} = 4.00$ $p < .001$), but non-significant for the Chinese guise ($p > .05$). The detriment of lexical/typographical errors were similar for the three guises (ENG: $M_{difference} = 1.52$, SWE: $M_{difference} = 1.77$, CHI: $M_{difference} = 1.53$), with significant decrease in the judgements for the error texts in comparison to the control texts (ENG: $F_{(3)} = 16.07$, $p < 0.001$; Tukey HSD of $p < 0.01$ for Grammatical Control vs. Grammatical Error and Lexical Control vs. Lexical Error; SWE: $F_{(3)} = 11.80$, $p < 0.001$; Tukey HSD of $p < 0.01$ for Lexical Control vs. Lexical Error; CHI: $F_{(3)} = 9.04$, $p < 0.001$; Tukey HSD of $p < 0.01$ for Lexical Control vs. Lexical Error).

Finally, to investigate if the abovementioned scores varied to any substantial degree between the two formality conditions (i.e., differences between the results obtained in Study 2 and Study 3), statistical comparisons of mean score differences – as those reported above – were performed between equivalent formal and informal results (e.g., CHI formal grammatical error scores vs. CHI informal grammatical error scores, and so on). No significant differences were observed for any iteration, thereby suggesting that the difference in the intended formality of the texts did not have a noticeable bearing on the ratings given by the judges.

6. General discussion

We investigated whether native English speakers judge grammatical and/or lexical/typographical errors in texts with different levels of detriment depending on the writer's expected English language ability based on their ethnolinguistic group membership.

We found that in both formal and informal contexts, the error text by the authors from the higher language ability groups (ENG, SWE) were judged similarly - lexical and grammatical errors in both guises resulted in significant detriment in the ratings of English language ability. On the other hand, grammatical errors produced by the author representing the lower English language ability group (CHI) did not significantly affect the ability ratings. Our study suggests that while lexical errors in different types of texts create significant detriment in judgements of language ability for non-native and native speakers, judgement of ability based on grammatical errors are affected by the author's linguistic background. These results are consistent with the previous reports of sympathetic native speakers (e.g., Galloway, 1980) whereby native speakers appear to lower the linguistic bar for low ability speakers. In addition, our study suggests that the sympathetic native speaker effects do not surface only as a result of real-life exposure to the speaker's language abilities, but linguistic group membership (English vs. Swedish vs. Chinese) as indicated by an author's name affect the levels of sympathy demonstrated.

6.1. Why the different error judgements for Swedish/English and Chinese speakers?

We suggest that the different judgments between the texts produced by the Swedish/English and Chinese guise was created by a sympathetic native speaker effect which was influenced by the writer's expected English language ability. This could be at least partly driven by the judges' assumption that the relatively subtle grammatical errors in the Swedish and English guises' texts were slips – thus indicating sloppiness – while the same grammatical errors might have been assumed to have surfaced in the Chinese guise's text due to lack of linguistic knowledge and thus in a way out of the writer's control. Alternatively, this might have arisen as a consequence of the social comparison bias (Garcia, Song & Tesser, 2010) which might have affected the judges' decision-making process. Namely, individuals who

saw the high ability authors as possible threats to their own native level of English writing, might have judged the ENG author (a perceived direct threat) and SWE author (also a perceived threat, but through high societal opinion of Swedish people, culture and English language abilities) more critically. Harsher judgements through perceived competition for English language ability would support findings of previous studies (Garcia, Song & Tesser, 2010; Huiyuan et al., 2016) which found evidence of social comparison bias in their employee recommendation-based experiments. The effect would not influence judgements on the CHI author, as a perceived lower language ability (as per Study 1) reduces the possibility of competition in judges' eyes, thus enabling more influence from the sympathetic native speaker effect.

Alongside these suggestions, it is useful to consider if anything else could have brought about the observed effects. First, even though the judges were told that their responses would be analysed anonymously and that they should give as accurate judgements as they could, the judges might have felt more at ease being critical towards the English and the Swedish guise, than Chinese guise that represented a different ethnic background. However, our study did not elicit judgements relating to the guises' personality traits (e.g. trustworthiness, friendliness) or social ranking, but asked the judges to give their perception of the guise's English language level. Furthermore, we did not find a difference between the three groups in the control or lexical/typographical texts (but only when grammatical errors were present). Thus, it is unlikely that the different ethnic guises impacted the results from the political correctness viewpoint.

Second, we found that the expected language level resulted in differences in the language ability judgements between the two non-native (SWE, CHI) guises, and that the Swedish guise was judged similarly to the native speaker guise (ENG). However, it is possible that the questions used in the study had different meanings when referring to a native

versus a non-L1 speaker, in particular when it came to the two questions eliciting judgements for overall proficiency (Command of English; Naturalness). This is because native speakers might be expected to have a high base level of language command and naturalness while for non-L1 speakers the expected base level might be lower. To investigate the possibility that these types of differences in the meaning of the questions used could have affected our results, we performed the same analyses as reported in Study 2 and Study 3 but analysed the overall proficiency (command, naturalness) and lexicogrammar (grammar, words, spelling) questions separately. We found that this did not change our results (results tables can be found at https://osf.io/gn2pr/). Even though this extra analysis strengthens our argument for the language level expectation affecting language ability judgements, this does not completely take away the possibility that the questions meant slightly different things when considering a native vs. a non-L1 English speaker.

Third, even though we used the Census data (Office for National Statistics, 2016) and conducted a survey on native speakers' expectations on non-L1 speakers' language ability, it is possible that our judges' expectations for Swedish and Chinese speakers did not accurately match our assumptions. The expected language level on the level of sympathy native speakers project could be further investigated in experimental contexts where the expectation can be manipulated. Furthermore, it would be beneficial to investigate in the future if the same effects as we found in our survey study would also surface in more authentic contexts.

6.2. The effect of different error types

We found that, similarly to previous studies (e.g., Brandenburg, 2015; Queen & Boland, 2015) lexical/typographical errors seemed to be relatively bothersome to the judges, given that, for all three guises, these types of errors had a significant negative effect on the language ability ratings. Why is it that lexical/typographical errors are judged more harshly

than grammar-related errors? One possibility is that speakers are aware of the fact that learning words and the correct spelling is something that rests upon conscious, declarative processes in native and non-L1 speakers, and thus something that anyone should/could learn in an appropriate learning environment – even non-L1 speakers who have learned their L2 after infancy. Possibly because of this, typographical errors are often perceived as laziness on behalf of the writer to review their draft work (Figueredo & Varnhagen, 2005). Grammar on the other hand might be viewed more of an automatic or unconscious, procedural process which is more difficult to overtly control or improve and is more susceptible to age of acquisition effects in non-L1 speakers. Thus, grammatical errors particularly in non-L1 speakers might be perceived to be more out of the speaker's control and thus, are not viewed as negatively in low level speakers than lexical/typographical errors.

6.3. Detriment of errors in formal and informal texts

We found that the judgements were very similar between formal (Study 2) and informal (Study 3) contexts qualitatively and quantitatively. It appears that this finding contradicts some of the earlier studies in which judgements have been found to be more lenient in informal than formal contexts (Cargile, 1997; Mozafari, et al., 2019). It is important to note that errors in both formal and informal context have been found to result in negative outcomes rather than errors being ignored altogether in informal contexts - it is just the level of detriment that has been found to be different (Derwing et al., 2002; Mozafari et al., 2019; Queen and Boland, 2015). Our results are in line with the reports that errors in all contexts affect judgement, it is just the difference in the detriment between the two contexts that we failed to find. The following four suggestions can explain this. First, it might be that even though our judges read texts differing in the level of formality (a cover letter or Facebook posts), the context of the Facebook posts (on paper) did not sufficiently mimic real life

Facebook posts (online) and thus were considered more formal than in more naturalistic contexts.

Second, as far as we are aware, our study is the first that has investigated the difference between formal and informal contexts by using the exact same errors in the two contexts and matched the length of the texts, making the formal and informal contexts more comparable than some of the previous studies (e.g., Cargile, 1997). It is possible that some of the differences between formal and informal contexts observed in previous studies were brought about by the difference in the errors/texts compared.

Third, the context of testing (e.g. out in the community vs. in the classroom) between our study and some previous studies might have affected the results found in different studies.

Fourth, we elicited language ability judgements by using a written guise technique, while most previous studies investigating perceived language ability in formal versus informal contexts have used a different method (e.g., Derwing et al., 2002). These differences in methods used are likely to have contributed to the differences in the results found.

6.4 Implications of the current study in L2 pedagogical settings

Our study has two key implications that should be considered in L2 pedagogical settings. First, similarly to Bradenburg (2015), Queen and Boland (2015) and Santos (1988) lexical/typographical errors were found to be detrimental to non-L1 writer's language ability judgements in formal and informal contexts regardless of the writer's linguistic background. This result indicates that focus on the accurate spelling (e.g, *their* for *there*) has to be emphasised, in particular, if the writer's aim is to indicate high language ability, which needs to be kept in mind even when using communicative pedagogical approaches that focus on comprehensibility over accuracy.

Second, importantly yet somewhat disappointingly to non-L1 speakers from high ability groups, our results indicate that a linguistic group's collective language level has an inverse effect on judgements of language ability based on errors of grammar in tests. More specifically, the writer's perceived language proficiency level will take a bigger hit if a writer from a high ability non-L1 speaker group produces errors of grammar in comparison to a writer from a lower ability non-L1 group producing the same errors. Given that our judges were not educators, but members of the general native speaker public, these effects are likely to take place outside the classroom in the 'real word' in a number of contexts. Thus, even though grammar is only a part of language and often contributes to a lesser degree to speakers meeting their communicative goals than lexical items, the effect of grammatical errors in high ability groups needs to be kept in mind in particular in contexts where the writer wants to express linguistic ability and the reader cannot base their expectation on actual evidence of the speaker's language level (e.g. a job cover letter or an online dating site). Thus, in pedagogical contexts, while meaning centred approaches might be a good choice with some learner groups, emphasis on accurate grammar might be needed in particular for speakers from high ability ethnolinguistic groups.

While noting points to consider in L2 writing teaching and learning, it remains that L2 language learners and their writing will continue to be affected by native speaker preconceptions, due to what appears to be a broad sociocultural issue. It is therefore suggested, albeit beyond the remit of this paper, that a rethink in broader ideology of non-L1 speakers is needed. This may lead to the interactions with those in the fields of intercultural communication and educational planners working with the national curriculum in order to address this issue.

7. Conclusion

Judgements of language competency are affected by various factors. The current study investigated one such factor: expectation-related sympathetic native speaker effects for grammatical and lexical/typographical errors in written texts between native (i.e., English) and two non-L1 English speaker guises (i.e., Swedish and Chinese). Through two text-based written-guise studies, it was found that the linguistic group membership the guise affected the grammatical ability judgements given. This suggests that sympathetic native speaker effects surface not only based on actual evidence of the speaker's language ability, but also due to language expectation at group level. Further study into this area with alternative or additional guises, materials, languages and methods is highly encouraged.

References

Ambridge, B., et al. (2020). The crosslinguistic acquisition of sentence structure:
Computational modeling and grammaticality judgements from adult and child speakers
of English, Japanese, Hindi, Hebrew, and K'iche'. *Cognition*, *202*, 104310

Appelman, A., & Schmierbach, M. (2018). Make no mistake? Exploring cognitive and
perceptual effects of grammatical errors in news articles. *Journalism & Mass
Communication Quarterly*, *95*, 930-947.

Ayhan, Ü. & Uğur Türkyılmaz, M. (2015). Key of language assessment: Rubrics and rubric
design. International Journal of Language and Linguistics, 2, 82-92.

Beason, L. (2001). Ethos and error: How business people react to errors. *College
Composition and Communication*, *53,* 33–64.

Bender, E. M. (2005). On the boundaries of linguistic competence: Matched-guise
experiments as evidence of knowledge of grammar. *Lingua*, *115*, 1579–1598.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child development, 57*,
498-510.

Birkelund, G. E., Rogstad, J., Heggebø, K., Aspøy, T. M. and Bjelland, H. F. (2014).
'Diskriminering i arbeidslivet – Resultater fra randomiserte felteksperiment i Oslo,
Stavanger, Bergen og Trondheim', *Sosiologisk tidsskrift*, *22,* 352–82.

Brandenburg, L. C. (2015). Testing the recognition and perception of errors in context.
Business and Professional Communication, *78*, 74–93.

Cargile, A. C. (1997). Attitudes towards Chinese-accented speech: An investigation in two
contexts. *Journal of Language and Social Psychology*, *16*, 434–444.

Carpusor, A. G. & Loges, W. E. (2006). Rental discrimination and ethinicity in names.
Journal of Applied Social Psychology, 36, 934-952

Cotton, J., O'Neill, B. S., & Griffin, A. (2008). The "Name Game": Affective and hiring

reactions to first names. *Journal of Managerial Psychology,* 23, 18-39

Council of Europe. (2020). *Common European Framework for Reference for Languages: Learning, teaching, assessment - Companion volume*. Council of Europe Publishing, Strasbourg.

Dąbrowska, E. & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*, 604–615.

Davies, E. E. (1983). Error evaluation: The importance of viewpoint. *ELT Journal*, *37*, 304–311.

Delamere, T. (1996). The importance of interlanguage errors with respect to stereotyping by native speakers in their judgements of second language learners' performance. *System*, *24*, 279–297.

Derwing, T. M., Rossiter, M. J., & Ehrensberger-Dow, M. (2002). "They speaked and wrote real good": Judgements on non-native and native grammar. *Language Awareness*, *11*, 84–99.

Dragojevic, M., Fasoli, F., Cramer, J., & Rakić, T. (2021). Toward a Century of Language Attitudes Research: Looking Back and Moving Forward. *Journal of Language and Social Psychology*, *40*, 60–79.

Dragojevic, M., Giles, H., Beck, A. C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication monographs*, 84(3), 385-405.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2020). *English Language*. Ethnologue: Languages of the World.

Fayer, J. M. & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, *37*, 313-326.

Figueredo, L. & Varnhagen, C. K. (2005). Didn't you run the spell checker? Effects of type

    of spelling errors and use of a spell checker on perceptions of the author. *Reading*

    Psychology, *26*, 441–458.

Foster, J. (2007). Treebanks gone bad: Parser eveluation and retraining using a treebank of

    ungrammatical sentences. *International Journal of Research and Analytical Reviews*,

    10, 129-145.

Galloway, V. B. (1980). Perceptions of the Communicative Efforts of American Students of

    Spanish. *Modern Language Journal*, *64*, 428–433.

Garcia, S., Song, H. & Tesser, A. (2010) Tained recommendations: The social comparison

    bias. *Organizational Behaviour and Human Decision Processes. 113*, 97-101.

Garrett, P. (2010). *Attitudes to language*. Cambridge University Press.

Garrett, P., Coupland, N., & William, A. (2003). *Investigating language attitudes: Social*

    meanings of dialect, ethnicity and performance. University of Wales Press.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*, 211–227.

Greenwald, A., McGhee D. E., & Schwartz, J. L. K. (1998). Measuring individual

    defferences in implicit cognition: The implicit association test. *Journal of Personality*

    and Social Psychology, *74*, 1464-1480

Hairston, M. (1984). Not all errors are created equal: Nonacademic readers in the professions

    repond to lapses in usage. *College English*, *43*, 794–806.

Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning:

    The influence of maturational state on the acquisition of English as a second language.

    Cognitive Psychology, *21*, 60–99.

Kantz, M. & Yates, R. (1994). Whose judgments? A survey of faculty responses to common

    and highly irritating writing errors. In *Partial Proceedings of the Annual Conference of*

    the Assembly for the Teaching of English Grammar, 38–54.

Klimanova, L. & Dembrovskaya, S. (2013). L2 identity, discourse and social networking in Russian. *Language Learning and Technology, 17,* 69-88.

Kreiner, D. S., Schnakenberg, S. D., Green, A. G., Costello, M. J. & and McClin, A. F. (2002). Effects of spelling errors on the perception of writers. *Journal of General Psychology*, *129*, 5–17.

Lambert, W. E., Hodgson, R.C., Gardner, R. C. & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*, 44–51.

Leonard, D. J. & Gilsdorf, J. W. (1991). Language in change: Academics' and executives' perceptions of usage errors. *The Journal of Business Communication*, *27*, 137–158.

Levon, E., Sharma, D., Watt, D. J. L., Cardoso, A., & Ye, Y. (2021). Accent bias and perceptions of professional competence in England. *Journal of English Linguistics*, 49, 355-388. https://doi.org/10.1177/00754242211046316

Mirshahidi, S. (2016). I find you attractice but I don't trust you: the case of language attitudes in Iran. *Journal of Multilingual and Multicultural Development*, *38*, 146-159.

Mozafari, A., El-Alayli, A., Kunemund, A. & Fry, T. (2019). Impressions of businesses with language errors in print advertising: Do spelling and grammar influence the inclination to use a business? *Current Psychology*, *38*, 1721–1727.

Mumford, S. (2009). An analysis of spoken grammar: The case for production. *ELT Journal*, *63*, 137-144.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *63*, 59-82

Office for National Statistics ; National Records of Scotland ; Northern Ireland Statistics and Research Agency (2016): 2011 Census aggregate data. UK Data Service (Edition: June

2016). DOI: http://dx.doi.org/10.5257/census/aggregate-2011-1

Planken, B., van Meurs, F. & Maria, K. (2019). Do errors matter? The effects of actual and perceived L2 English errors in writing on native and non-native English speakers' evaluations of the text, the writer and the persuasiveness of the text. *International Journal of English Language Teaching*, *6*, 1–16.

Preston, D. R. (1981). The Ethnography of TESOL. *TESOL Quarterly*. *15,* 105-116.

Queen, R. & Boland, J. E. (2015). I think your going to like me: Exploring the role of errors in email messages on assessments of potential housemates. *Linguistics Vanguard*, *1*, 283–293.

Roberts, F. & Cimasko, T. (2008). Evaluating ESL: Making sense of university professors' responses to second language writing. *Journal of Second Language Writing, 17*, 125-143.

Rubin, D. L., & Williams-James, M. (1997). The impact of writer nationality on mainstream teacher's judgments of composition quality. *Journal of Second Language Writing*, *6*, 139–153.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, *22*, 69-90.

Schmid, M. S., Gilbers, S. & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch-English bilingualism. *Second Language Research*, *30*, 129–157.

Schmitt, N. (1993). Comparing native and nonnative teachers' evaluations of error seriousness. *Japanese Association of Language Teaching Journal*, *15*, 181–191.

Spillner, B. (2017). *Error analysis in the world*. Frank and Timme GmbH.

Tajfel, H. & Turner, J. C. (2004). The social identity theory of intergroup behaviour. Key readings in social psychology. In John T. Jost & Jim Sidanius (Eds.), *Political*

psychology: Key readings, Key readings in social psychology (pp. 276–293).

Psychology Press.

The American Council on the Teaching of Foreign Languages. (2015). *Performance*

descriptors for language learners. ACTFL Publishing, Alexandria.

Wheeler, R., Cartwright, K. B., Swords, R. (2012). Factoring AAVE into reading assessment

and instruction. *The Reading Teacher*, *65*, 416–425.

Wilcox, K. C., Yagelski, R., & Yu, F. (2014). The nature of error in adolescent student

writing. *Reading and Writing*, 27, 1073-1094.

---

[1] We use the term *non-L1* to refer to second and foreign languages, as well as to a non-balanced or sequential bilingual's weaker language(s).

[2] We use the term 'Chinese' to refer collectively to Cantonese, Mandarin and 'all other Chinese' speakers. However, we do not mean to imply that the different variants of Chinese would be the same or that the people speaking these variants are a homogenous group.

[3] Term used as per UK Government census publications.

[4] Note also that in addition to the permanent residents, whose data is included in the census, there are also short-term international students, tourists and other visitors in the UK whose language use can impact British people's perception of speaker groups' language ability.

[5] We use the term lexical/typographical errors as it is not always clear whether the errors we included in this category instantiate lexical or typographical errors.

[6] All data and materials used are available at https://osf.io/gn2pr/

[7] English – first name: https://tinyurl.com/yy3zr5qx; surname: Prof Richard Coates, personal communication, July 2019.

Swedish – first name: https://hejsweden.com/en/popular-swedish-girl-and-boy-names/; surname: https://www.nordicnames.de/wiki/Main_Page

Chinese – https://www.thechairmansbao.com/the-ten-most-common-names-in-china/